DOCUMENT RESUME

ED 173 173 **SE 028 441**

AUTHOR	Wagner, William J.
TITLE	Application of Stochastic Learning Theory to Elementary Arithmetic Exercises. Technical Report No. 302. Psychology and Education Series.
INSTITUTION	Stanford Univ., Calif. Inst. for Mathematical Studies in Social Science.
REPORT NO	SU-IMSSS-TR-302
PUB DATE	4 Jun 79
NOTE	131p.; Not available in hard copy due to marginal legibility of original document
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Educational Psychology; Elementary Education; *Elementary School Mathematics; *Learning Theories; Mathematical Models; Mathematics Curriculum; *Mathematics Education; *Mathematics Instruction; Prediction; *Research

ABSTRACT

The application of a linear learning model, which combines learning theory with a structural analysis of the exercises given to students, to an elementary mathematics curriculum is examined. Elementary arithmetic items taken by about 100 second-grade students on 26 weekly tests form the data base. Weekly predictions of group performance on addition items are made, based upon earlier observations. The model is tested by studying the accuracy of the predictions, by varying parameters of the model, and by comparing its predictions with those made from a regression analysis. The study indicates that the most important skill to consider in making predictions of performance on addition items is carrying. The skill of adding two digits made a much smaller contribution. (MP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED173173

APPLICATION OF STOCHASTIC LEARNING THEORY TO ELEMENTARY ARITHMETIC EXERCISES

by

Willian J. Wagner

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Marquerite K. Shaw

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TECHNICAL REPORT NO. 302

June 4, 1979

PSYCHOLOGY AND EDUCATION SERIES

Reproduction in Whole or in Part Is Permitted for

Any Purpose of the United States Government

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA 94305

SE 028 441

ACKNOWLEDGEMENTS

I wish to express sincere thanks to Dr. Barbara Searle, Director of the Nicaragua Radio Mathematics Project, for her assistance, advice, and warm personal support through all the stages of this dissertation.

Professor Patrick Suppes, Director of the Institute for Mathematical Studies in the Social Sciences, and the principal advisor for this dissertation, provided not only a stimulating working environment but also useful ideas, helpful criticism, and continuing support throughout the project. Mario Zanotti of the Institute provided valuable advice on statistical matters.

Finally, to my family go my love and fondest thanks: To my parents, who kept the dream alive, to my wife Linda, who made it possible, and to my children, Alice and Mark, who made it interesting.

TABLE OF CONTENTS

	Page
I. Introduction	1
II. Curriculum and Learning Theory	5
1. Learning theory	5
A. Incorporating Stochastic Theory	6
B. Further Developments	11
C. Experimental Studies	12
2. Learning Theory and Arithmetic	14
3. Structural analysis of exercises - performance models	17
4. Applying the models to a curriculum	21
III. The Scope of this Study	25
1. The curriculum	25
2. The weekly testing program	27
3. A performance model for addition	30
4. Introducing learning into the model	32
5. Estimating the parameters	35
A. Initial Skill Levels	35
B. The Learning Parameter	37
6. Judging the model	39
7. Summary of the analysis of the model	42
8. Modifications of the model	46
A. Analysis by class	46

TABLE OF CONTENTS (continued)

	Page
IV. Explicit Formulation and Results	48
1. Notation	48
2. Analysis of exercises	50
3. Estimation of the parameters	55
4. Estimating θ for a week	63
5. Using the model - predicting performance on future exercises	69
A. "Next week" predictions	69
B. "Far future" predictions	74
C. Effect of varying the initial skill parameters	76
6. Predictions using modifications of the model	78
A. Using Fewer Parameters	78
B. Constant Learning Parameter	81
C. Analysis by Class	83
7. Comparison with regression predictions	88
V. Summary and Discussion	99
VI. Suggestions for Future Research	103
VII. References	107
VIII. Appendices	
A. Sample student answer sheets	109
B. The MLE estimate across exercises on a given week	113

LIST OF TABLES

	Page
1. Schedule of Testing for Each Class	28
2. Description of Vertical Addition Objectives	52
3. Vertical addition Objectives Included in the First 5 Weeks of Testing	54
4. Estimating θ for Each Exercise	57
5. Estimates of θ for Each Exercise	61
6. Values of Log (L) Vs. θ	66
7. Comparison of θ which Maximized the Likelihood for a Week (Observed MLE), and the Average of the MLE Estimates for a Given Week (Predicted MLE)	67
8. Effects on 'Next Week Predictions' of Variations in Initial Skill Parameters	77
9. Results of Predictions Using Two Skill Parameters	79
10. Results of Predictions Using One Skill Parameter	82
11. Predictions Using a Constant Learning Parameter	84
12. Estimates of Parameters for Individual Classes	86
13. Itemized Predictions for Two Classes	87
14. Summary of Predictions for Each Class	88
15. Summary of Results of Modified Class Analyses	90

LIST OF TABLES (continued)

	Page
16. Regression of Observed Proportion Correct on Structural Features of Exercises	94
17. Comparison of Predictions by Regression and Predictions by Model	95
18. Summary of Results	100

LIST OF FIGURES

	Page
1. Typical Sequences in Mathematical Learning Theory	23
2. Estimates of θ for each Exercise vs. Week of Estimation	59
3. Weekly Estimates of θ (Averaged Across Exercises) vs. Week of Estimation	68
4. 'Next Week Predictions' vs. Observed Proportion Correct	72
5. Predicted Minus Observed vs. Week	73
6. 'Far Future Predictions' - Prediction of Proportion Correct for Five Exercises, vs. Week of Prediction	75
7. Regression Predictions vs. Observed Proportion Correct	97
8. Regression Residuals (Transformed) vs. Week of Prediction	98

LIST OF APPENDICES

	Page
A. Sample Student Answer Sheets	109
B. The Likelihood Estimate for a Week	113

I. INTRODUCTION

During this century mathematical learning theory has inspired more hopes among researchers and practitioners than it could ever satisfy. The elementary, intuitively pleasing axioms and the richness of their mathematical consequences have led to a large body of theoretical work and to extensive experimentation with simple task learning. Throughout this time, however, there have been few, if any, attempts to apply the theory to actual classroom data, in part because the models quickly become too complex or mathematically intractable as the learning tasks more closely resemble those encountered in schools.

The complexity of even the simplest arithmetic exercise led Suppes and his co-workers to turn their attention to the structure of exercises. The performance models which they developed ignored learning over time, but produced regression predictions which were remarkably good (Suppes, Hyman and Jerman, 1967; Suppes and Morningstar, 1972).

Mathematical learning theory has not been applied to arithmetic exercises taken by students in actual classroom settings. A necessary condition for such applications is a set of well-designed exercises presented in a consistent form over significant time. In the context of the Nicaragua Radio Mathematics Program, which has been designed and administered for USAID by the Institute for Mathematical Studies in the Social Sciences, a body of such data has recently been gathered. Although collected to be part of the formative evaluation of the

curriculum, the tests are carefully designed and the results lend themselves to to a more theoretical situation as well.

This dissertation involves the application of a linear learning model to an elementary mathematics curriculum. The model combines learning theory with a structural analysis of the exercises given to students. Weekly predictions of group performance on addition items are made, based upon earlier observations. The model is tested by studying the accuracy of the predictions, by varying various parameters or structure of the model, and by comparing its predictions with those made from a regression analysis similar to that found in the work cited above.

Elementary arithmetic items taken by about 100 second-grade students on 26 weekly tests form the data base. All data analysis was performed on the PDP-10 computer at the Institute for the Mathematical Study of the Social Sciences, at Stanford University. With the exception of the regressions, all programs were written in the language SAIL by the author. Regressions were done with the STP programs (University of Western Michigan Statistical Package).

Section II, which follows this Introduction, is a general discussion of stochastic learning theory and its application to a well defined curriculum. Section III describes the particular curriculum and testing program involved, and gives the details of the model which is developed and the means of studying it. Section IV explicitly delineates the analysis and testing of the model, including the results

obtained. A discussion of the results concludes the main text.

This dissertation depends heavily upon established theories and existing data. Its significance and contribution depend primarily on the following considerations:

- Learning theories have been rarely tested with classroom test data, particularly for a period as long as a full school year.

- The integration of a performance model for elementary arithmetic with stochastic learning theory is unique.

- When test data are collected in the context of a coherent and complete curriculum like this one, certain simplifying assumptions are possible which make the mathematics of the analysis much more tractable than is usually the case in stochastic learning theory.

- An important aspect of the proposed analysis is that it is predictive rather than explanatory; that is, one subset of the data is used to predict proportion correct for exercises in a different subset. The model which is tested could be used to predict future performance of a group, given past results.

- The analysis requires derivation of an original expression for the maximum likelihood estimator of the learning

parameter.

Finally, in addition to certain theoretical contributions discussed above, the essential practicality of this study must be emphasized. The curriculum is Nicaraguan, but is little different from any in the Western world. The arithmetic is elementary, but is studied throughout the world in much the same format; and similar content makes up a large proportion of the mathematics of the elementary school years. Thus any contributions which the proposed study can make to the understanding of teaching and learning in this area can have an impact beyond the particular location, curriculum, and grade level which is studied here.

II. CURRICULUM AND LEARNING THEORY

II-1. Learning Theory

Learning theories abound, but several common threads exist. Usually there is reference to a stimulus and a response, to a connection between them, and a suggested way of predicting and describing the improvement or acquisition of a skill. One way to classify learning theories is according to whether they tend to be quantitative or qualitative. In qualitative theories descriptive relationships are postulated between stimuli and responses, and learning is said to be influenced by these relationships in some way which makes intuitive sense. There is much speculative imagination to be found here. Thorndike spoke of "bonds" and "transfer", and Guthrie of "association by contiguity". A modern counterpart of them is Bruner, whose "theorems of instruction" are prescriptions for maximizing learning which might be drawn, in the tradition of E. L. Thorndike, from experience or empirical research, but which are not theorems in any strict sense of the word (Bruner, 1964).

Quantitative theories of learning have been dated to 1907. (Coombs, Dawes and Tversky, 1970, Estes, 1962, and Bush, 1960 provide summaries of these early efforts). The two approaches have continued to develop in a parallel manner with similar goals, the differences reflecting the variety of training and background among psychologists. Thus while E. L. Thorndike observed the effects of practice and the limitations of transfer, and then produced his "laws of learning", his

contemporaries, Thurstone, Hull, and Gulliksen sought laws or sets of axioms as an intermediate step toward formulas which would fit similar observations. In 1934, for example, Gulliksen essentially expressed Thorndike's "law of effect" in mathematical form. (Coombs, et al, 1970).

A. Incorporation of stochastic theory

Quantitative approaches to learning theory began to flourish in the 1950's with the introduction of stochastic relationships in the work of Estes (1950) and Bush and Mosteller (1955). They defined learning in terms of the probability of a correct response being given, and attempted to ground the formulas with rigorous axioms and definitions. Although they began with different fundamental ideas, both approaches lead to a similar mathematical formulation, the linear learning model which forms the basis of this dissertation.

Each of these stochastic approaches conceives of learning in the context of a sequence of trials which consist of the presentation of a stimulus which is followed by a response, an outcome (e.g., reinforcement or not), and a change in the distribution of the probabilities related to each of the possible responses. The differences are in the assumptions concerning the nature of the stimuli, and in the way the probabilities change with each trial.

Estes' conception of "stimulus sampling theory" follows explicitly in the spirit of Thorndike's bonds and Guthrie's stimulus-response association. He considered complex stimuli each made

up of stimulus 'elements', different combinations of which produce a whole spectrum of stimuli, in the way a lexicon is built from an alphabet. Each stimulus is followed by a response. If the response is 'reinforced', then each of the elements of the stimulus are said to be 'conditioned' to that response. The central axiom which Estes applied to this structure was that the probability of a response to a given stimulus is equal to the proportion of the elements of that stimulus which are conditioned to the response.

Adding a few assumptions to this structure, Estes was able to derive a result whose mathematical simplicity has made it quite useful. The theorem is that, when averaged across individuals, the probability of a response changes in a linear fashion if reinforcement occurs. The linear relationship is presented below in two algebraically equivalent forms:

$$p_{n+1} - p_n = \theta(1 - p_n)$$

OR (Equation 1)

$$p_{n+1} = (1 - \theta)\, p_n + \theta$$

Here p_{n+1} is the probability of a response R on trial n+1, given that R was reinforced on trial n. In Estes' formulation θ is a constant equal to the probability that any given stimulus element is present on trial n. The first form of equation 1 states that the increase in the probability of R is directly proportional to the

difference between the upper limit of the probability (in this case 1.0) and the current value. Thus the value of p_n will increase with n (if there is reinforcement) but the graph of p_n will be concave down because of the continual decrease in the magnitude of the increment as the limiting value is approached. (The original attraction of the linear model was based upon the shape of this predicted learning curve, for it had commonly been observed in various types of learning situations.) The second form of equation 1 explicitly gives the linear transformation which is applied to p_n to produce p_{n+1} when there is reinforcement of R.

At about the same time, Bush and Mosteller developed the so-called "operator" model from three assumptions: that the probability of a given incorrect response is independent of the set of incorrect responses, that the response probabilities vary in a manner independent of all but the most recent trial, and that there exists an operator defining the transformation on the response probability with reinforcement. From these they derive Estes' linear relation for an individual's response probability (equation 1).

The differences in the models need to be emphasized. In stimulus sampling theory the set of stimuli, and therefore the set of possible probabilities is finite; thus the door is opened to the application of the theory of finite state Markov processes, which has proved quite fruitful. (The operator model assumes a continuous distribution of probabilities.) The axioms of stimulus sampling theory refer to more basic mental processes and thus can be related more easily

to other work in psychology. Another important difference is in the basic conception of learning. Stimulus sampling theory assumes that learning is "all or none", since once a given set of stimuli is associated with a particular response through reinforcement, the theory predicts that the response will be repeated with probability 1 if the same stimulus set is repeated. In contrast, the operator model defines learning explicitly in terms of probabilities which change gradually according to equation 1. Finally it should be emphasized that stimulus sampling theory gives the linear relation as an average across individuals, and the operator model is in terms of the response probabilities of an individual.

Another significant approach of the 1950's, Luce's "beta" model (Luce, 1959), had some of the flavor of both these theories. He defined response probabilities to be indirectly related to outcomes of trials, the intermediate stage being linear changes in so-called 'response strengths', a concept with roots in classical learning theory. Assuming that the probability of a response is proportional to its strength relative to all possible responses, Luce showed that probabilities must then change according the following relation:

$$p_{n+1} = \frac{b\, p_n}{(1 - p_n) + b\, p_n}$$

where b (beta in Luce's paper) is dependent on the particular stimulus presented.

The complexity of this expression has led to difficulty in making predictions which can be checked by experimentation. However it has been compared with the linear model on at least two occasions. Luce (1959, p. 96-99) shows that with the addition of a few assumptions the beta model will approach the linear model. Also, Coombs, et al (1970, p. 275-6) report an experiment in which the recursive expressions of the two models fit the data equally well.

Thus we have three quite different conceptions of learning - stimulus sampling theory, the operator model, and Luce's 'beta' model - which all may be approximated in some way by a simple linear expression for change in response probability. The linear model therefore has two important advantages: it is consistent with various views of the fundamental nature of learning, and its simplicity allows it to be applied readily to various experimental situations to produce predictions which may be checked with data. It is not surprising, then, that the linear learning model was popular among researchers in the 1960's. Its attraction is well expressed by Estes:

> "The intuitive naturalness and simplicity of linear models are evidently exceedingly compelling considerations, for nearly every theorist to try his hand at the quantification of learning .... has assumed that learning as a function of successive rewards, punishments, or similar operations proceeds by linear transformations of response strength or response probability." (Estes, 1962, p. 110-111).

B. Further developments

Because of its more rigorous structure and closer relation to mental processes, stimulus sampling theory has proved to be rich in mathematical and psychological consequences. Estes and Suppes (1959) give a more general theoretical discussion of the model and its properties. They begin with the assumption of linearity and add an effect on p_n when R is not reinforced, and an effect on p_n when a different response is reinforced. The relevant results of Estes and Suppes' formulation are presented in less rigorous form here:

If response R was reinforced on trial n then:

$$p_{n+1} = (1-\theta)\, p_n + \theta \qquad \text{(Equation 2a)}$$

(same as Estes' derived result)

If a response different from R was reinforced on trial n then:

$$p_{n+1} = (1 - \theta)\, p_n \qquad \text{(Equation 2b)}$$

If no reinforcement occurred on trial n then:

$$p_{n+1} = p_n \qquad \text{(Equation 2c)}$$

(Here θ is merely a constant 'learning parameter' between 0 and 1, having no specific definition as in Estes' work.)

In a later paper Estes and Suppes (1974) give a rigorous

formulation of stimulus sampling theory and present a proof of the applicability of Markov processes to the model. The mathematics of Markov processes, with the incorporation of "stages" of learning, has recently replaced the simpler linear model in popularity among quantitative learning theorists. The linear model remains important, however, because of its simplicity and its roots in stimulus sampling theory.

C. Experimental studies

This section concludes with a discussion of the experimental studies of these stochastic theories. Without attending to the details of the large number of studies (for reviews of research see Estes, 1962, Coombs, et al, 1970 Greeno and Bjork, 1973, and Cotton, 1976), it is fair to make two general comments. First, the experimental situations have ranged from rats in a T-maze to humans doing simple asks such as binary choice discriminations, recognizing patterns, and learning to correctly match paired words. The common denominator here is the simplicity of the tasks; they represent real learning situations only indirectly. Simple theories and simple tests are useful in illuminating the best directions to pursue, but sooner or later a theory must confront real world data. This dissertation appears to be unique, therefore, in its attempt to apply stochastic learning theory to actual classroom data gathered over an entire year. The complexities involved are great (some are addressed in the next section) but the goal must be to contribute to our understanding of actual phenomena.

The second comment concerns the degree to which the theories have fit the data of these simple experiments. While classical tests of fit are not generally employed, it seems fair to say that the theories have been less than overwhelmingly successful, as have been the attempts to discriminate between the success of two competing models with the same data. This should be kept in mind when considering the fit which is obtained in this study of classroom learning.

II-2. Learning Theory and Arithmetic: The Contribution of E. L. Thorndike

The most important modern application of psychological principles to the learning of arithmetic was made by E. L. Thorndike and his fellow workers early in this century. Others contributed to the understanding various aspects of the problem, but Thorndike was a rare combination of teacher, textbook writer, researcher and theoretician, and his effect, both indirect and direct, on the learning and teaching of arithmetic, was widespread. (Joncich, 1968; Cronbach and Suppes, 1969)

Thorndike's laws of learning represented an approach which was radically different from the theory of 'mental discipline', the prevailing view of the mind in the 19th century. His laws stressed repetition of rewarded behavior rather than learning either by transfer, by imitation or by pure intellectual effort. Although based upon experiments with simple tasks, he applied the laws in a wide variety of situations, including the learning and teaching of arithmetic (Thorndike, 1922).

The laws of learning may be summarized as follows: The law of readiness postulates that certain bonds are ready to be formed and that their formation is satisfying to the organism. The law of exercise asserts that repeatedly making a connection (i.e., following a given stimulus with a specific response) strengthens the bond and makes the response more likely when the stimulus recurs. The law of effect states essentially that rewards act to strengthen the bonds.

Thorndike's research dealt with learning of simple tasks, and with the effects of drill on performance on rote arithmetic exercises. The 'laws of learning' were hardly grounded in research, but there was a tendency for Thorndike to refer to them as the 'facts of learning', or the 'facts which science has revealed'. For example, in the introduction to Psychology of Arithmetic (1922) he says:

> "We now understand that learning is essentially the formation of connections or bonds between situations and responses, that the satisfyness of the result is the chief source that forms them, and that habit rules in the realms of thought as truly and as fully as in the realm of action". (p. iv)

In addition to his Psychology of Arithmetic, which was written for teachers, Thorndike also influenced the teaching of arithmetic with an elementary textbook series which attempted to put the laws of learning into practice. According to a recent review of early American arithmetic texts (Wagner, 1975), the significant changes which may be observed in these texts, the first to be grounded in some kind of theoretical framework, may be summarized as follows:

- The aim of the curriculum is to teach the specific skills of arithmetic, rather than the more abstract goal of mental discipline.
- What is today called the spiral approach to teaching is employed: the books are not organized by topic but by difficulty of the concept.
- The content is divided into discrete blocks and ordered in

difficulty, with each transition made as easy as possible to traverse.

- The problems are realistic, the material not esoteric, and the language more appropriate for children.

Thus we have an early example of pedagogy founded on theoretical principals of learning. Thorndike's assuredness that his laws of learning were indeed facts, his own personal eminence, the apparent foundation of the theory in science, his extensive writing and contacts with teachers, and the ease with which the ideas could be grasped, all combined to influence educational practice. The fundamental way in which elementary mathematics was taught in the United States followed the lead of Thorndike and other Progressives until the 1950's, when a new paradigm for teaching of the subject was introduced.

Thorndike's laws of learning also formed part of the basis of the later work on mathematical learning theory, beginning, as was noted in the previous section with Gulliksen in 1934. Stimulus sampling theory has roots in the work of many predecessors, but it shares with Thorndike's laws the basic attention to stimulus, response, and reinforcement. Estes sought in 1950 to quantify the way we think about learning, and and in many people's minds this 'way' had been molded largely by Thorndike.

II-3. Structural Analysis of Exercises - Performance Models

Extensive laboratory studies have been carried out in connection with the linear learning model described in the previous section, but as Suppes, Hyman, and Jerman (1967) observe,

> "...most of these experimental situations are much simpler in structure than even the simplest corresponding parts of elementary mathematics." (p.160)

The difficulty with the linear learning model is that the stimuli in even elementary mathematics exercises are too varied and complex to be adequately represented by the theory. Therefore Suppes and his co-workers set out to determine some small number of 'factors' (using the term non-statistically) which could be used to characterize a problem. In the study cited above, for example, three factors were used to characterize elementary addition problems. The factors were: a measure of the length or complexity of the problem, the magnitude of the sum, and the magnitude of the smaller addend. Regression analyses were carried out using observed proportion correct for each exercise as the dependent variable, and the values for each exercise of the three factors as the independent variables. The resulting regression equation then allowed prediction of the proportion correct for an exercise, from its values on each factor.

The accuracy of these predictions were, in their words, "encouraging". The only summary statistic reported to describe the accuracy of the predictions is a chi-square, which is not used in this

dissertation. To anticipate the measures of accuracy used here, we may calculate two of them from the representative subset of the fourth grade addition results which is given in Table 3, p. 182. The average of the absolute deviations of observed and predicted proportions correct for these 38 exercises is 0.12, and the correlation between observed and predicted is 0.83

Regression on structural features of arithmetic exercises has also been employed by Cromer (1974) in a study of elementary multiplication. Cromer's study included the use of factor analysis and was able to explain from 60 to 80 percent of the variance in proportion correct using from 3 to 10 features.

But structural features of an exercise cannot lead to significant insights into how students do exercises, and are thus of limited theoretical usefulness. To confront this inadequacy subsequent effort by Suppes was directed toward producing models which simulate the behavior of a person performing the steps involved in an exercise in elementary arithmetic. These are referred to as 'performance models'.

The first analysis of performance models was carried out with data obtained from the computer-assisted instruction program at Stanford (Suppes and Morningstar, 1972). In this study, the structural analysis of the exercises proceeded according to operations which were required of the student, rather than the magnitude of the numbers or the complexity of the problem. Potential sources of error were identified,

and the probability of making each kind of error was entered as a parameter into the model. In the operation of addition, for example, the errors included not carrying correctly, not entering the correct sum for each column, and carrying from a column which does not actually require it. Each addition exercise was then classified according to how many columns it had, how many carries were involved, from which columns these carries derived, and how many digits there were in each column. A count was then made of the total number of each kind operation which might produce an error.

Suppes and Morningstar represent the process of carrying out independent operations within an elementary addition exercise with a two-state (carry or no carry) machine which reads and processes pairs of digits, making errors probabilistically. Thus they refer to this kind of performance model as a two-state probabilistic automaton model. The potential errors are assumed to be independent, and therefore the proportion correct is predicted by the product of the probabilities of performing each sub-operation correctly. The error probabilities are estimated by maximizing the likelihood of the set of observations, and the predictions made for each exercise type according to the number of potential errors involved. Although no single summative measure of the accuracy of these predictions is given by Suppes and Morningstar, it seems to be extremely high. For example, the average of the absolute deviations between predicted and observed proportion correct by the three parameter model for one set of 20 addition items was 0.05 (no correlation could

be calculated from the table. (Pretest, Table 5, p. 183).

To allow for increased generality in performance models and to include specific ideas about the psychology of perceptual processing, Suppes has recently introduced an even more general model. He uses the concept of register machines to represent in detail the actual sequence of primitive procedures followed in doing elementary addition. The model is described in detail in an appendix in Suppes (1973). No empirical study of this complex model has been reported, but work is currently being carried out in this area.

The model to be analyzed in detail in this study is similar to the automaton model. Three operations (potential sources of error) will be identified, and a linear learning model applied to each of the three error probabilities. Likelihood will enter, but in a different way. Also, no observation will be used in the generation of its own prediction, as is the case with regression and with the automaton model described above. This model is unique in its incorporation of learning theory into a performance model, and the analysis is unique within learning theory because the Nicaragua weekly test data provides actual classroom results which allow detailed study of the change in skills over time.

II-4. Applying the Models to a Curriculum

A curriculum is a sequence of instruction organized to teach a particular body of knowledge. An actually implemented curriculum provides a real world environment within which to study learning, for in principle we know what has been taught, and, again in principle, we can measure what has been learned. Furthermore, because a curriculum is a sequence, we can study this relationship over time.

A curriculum thus is worlds removed from traditional experimental environments encountered in learning theory, which involve simple tasks and carefully controlled situations. Laboratory studies are important for theory building, but eventually the bridge to practical instructional considerations must be erected so that learning theory may contribute in that area. By studying learning theory in this real arena we have the opportunity to make two symmetrical contributions: previous generalizations from the laboratory to, say, the classroom, may be checked directly, and learning theory may be used to obtain insight into practical matters of instruction.

Furthermore, a curriculum offers an extra dimension to the traditional stimulus-response-reinforcement sequence of learning theory, in which the learning is assumed to be a product of the particular sequence which is experienced by the subject. In an educational setting the stimulus to which a student's response is generally measured is a test exercise. The stimulus in this context is

thus distinct from the instruction intended to produce learning. (I do not here ignore the effect of practice on test-taking proficiency, but only assume that intervening instruction is more influential in raising test scores.)

Therefore we are able to make a most important assumption which greatly simplifies certain computations: We may assume that reinforcement of correct responses is a continually occurring phenomenon and thus in each interval between testing the probability of correctly performing an operation is incremented. The specific implications of this important assumption are discussed in more detail in a later section. It may be summarized here by saying that the mathematical learning theory sequence represented in Figure 1A may be altered in the context of a curriculum to that shown in Figure 1B.

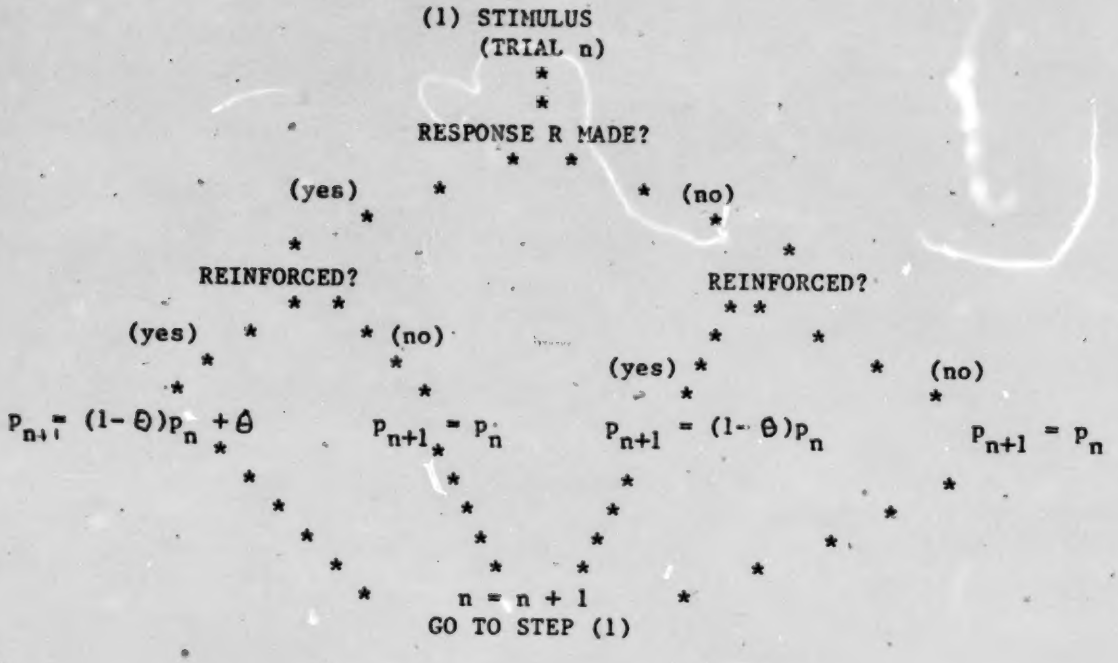
It must also be noted here that this assumption may be more suitable to some curricula than to others. For example, if the basic operation of carrying in addition is not taught or practiced after the first month, it makes little sense to increment the related probability automatically. Thus the simplifying assumption is more appropriate for a so-called spiral curriculum which provides for continuing cumulative review. We shall see that the Radio Mathematics Curriculum, to which learning theory and performance models have been applied in this study, does indeed satisfy this requirement.

Thus the nature of the curriculum may be useful in simplifying the analysis of the learning model. Also, the results of this analysis

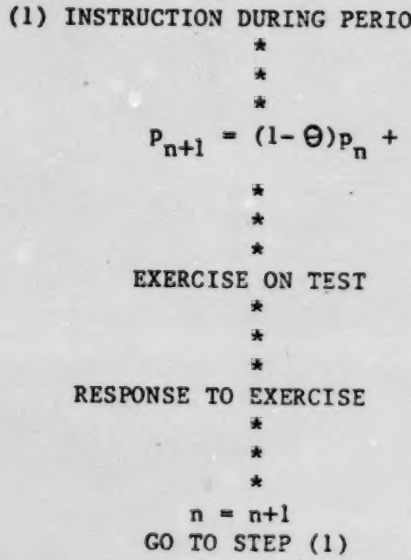
FIGURE 1 - TYPICAL SEQUENCES IN MATHEMATICAL LEARNING THEORY

(p_n = probability of response R on trial n)

A: AFTER ESTES AND SUPPES (1959)



B: MODIFICATION POSSIBLE WITHIN A CURRICULUM



may be useful to the designers and implementers of the curriculum, and to curriculum theorists generally. For example, if the model provides accurate predictions of student performance in future weeks, this information may be usable in the design of subsequent lessons. Furthermore, detailed post hoc analysis may be made of the relationships between actual instruction presented within a given period and the observed deviations from the predictions of the model, so that instructional strategies may be improved.

III. THE SCOPE OF THIS STUDY

III-1. The Curriculum

The Radio Mathematics Project is a program funded by USAID for the purpose of developing, implementing, and evaluating a mathematics curriculum for elementary schools which is broadcast over the radio in Nicaragua. Since its inception in 1973 the project has been under the direction of Ms. Jamesine Friend, Dr. Barbara Searle and Prof. Patrick Suppes of the Institute for Mathematical Studies of the Social Sciences at Stanford University. The design of the curriculum has been carried out in close cooperation with the Nicaragua Ministry of Education, whose formal mathematics curriculum guidelines have been followed throughout.

In the 1976 school year (February to November) the first full scale implementation without any local assistance from the Project was carried out. In 1977 115 schools are involved throughout the country in grades 2 and 3. Lessons are broadcast over the Nicaraguan National radio station at the same time each day for each grade. Instruction consists of a daily half-hour radio lesson and a postbroadcast lesson taught by the classroom teacher following a guide produced by the Project staff.

During the radio programs the students participate in a variety of activities, working exercises both orally and in their notebooks. Other activities such as games, songs and riddles change the pace of

the lessons and provide diversion from the mathematical activities. The radio lessons are characterized by a high rate of student response - averaging three responses per minute. The main characteristics of the radio lessons are the following:

a. A radio lesson consists of a number of independent segments whose main purpose is either instruction or entertainment.

b. All instructional segments and most entertainment segments require active responses from the students.

c. The instructional segments of a lesson are drawn from different topics and require different types of responses.

d. The instructional material takes the form of exercises supported by minimal amounts of explanation.

e. The correct answer for each exercise is given after the children have had the opportunity to respond.

More detailed accounts of the Project and the curriculum may be found in in Searle, et al (1976), and in Searle, Matthews, Suppes and Friend (1978), from which much of the above description was taken.

III-2. The Weekly Testing Program

During 1976, a weekly testing program was carried out in Grades 1 and 2 for the purpose of providing feedback to be used in the further development of the radio lessons. The study described here used the test results for Grade 2.

In designing this formative evaluation scheme for Grade 2, a stratified sample of 12 schools was selected from among the 20 schools receiving radio instruction in Masaya, the capital region. The sample was designed to be representative of all the schools receiving radio instruction, reflecting the distribution of average family income and the balance of urban vs. rural which were to be found among all the participating schools throughout the country. Beginning with the fifth week of the school year and continuing through the thirtieth, the same class in each school was tested at varying intervals. Each week from 4 to 6 classes were tested, each class being tested 11 to 13 times throughout the year. Approximately 100 students were tested each week. The actual schedule of testing is given in Table 1, and Appendix A contains a sample student answer sheet.

The tests, like the curriculum, were carefully designed. The second-grade curriculum is divided into 270 different objectives, and the test exercises are classified according to the objectives they are intended to assess. Objectives were tested throughout the year, some even being assessed before they were introduced in the lessons.

Table 1

Schedule of Testing for Each Class

	Class number											
Week	01	10	16	20	24	35	36	39	43	52	72	91
5			XX					XX		XX		XX
6	XX	XX					XX				XX	
7			XX	XX	XX	XX			XX			XX
8	XX	XX		XX			XX			XX	XX	
9			XX		XX	XX		XX	XX			XX
10	XX	XX		XX			XX			XX	XX	
11			XX		XX	XX		XX	XX			XX
12	XX	XX		XX			XX			XX	XX	
13			XX		XX	XX		XX	XX			XX
14	XX	XX		XX			XX			XX	XX	
15		XX	XX	XX	XX			XX			XX	
16				XX	XX			XX	XX	XX	XX	
17	XX				XX				XX	XX	XX	XX
18	XX					XX	XX		XX	XX		XX
19	XX	XX	XX			XX	XX					XX
20		XX	XX	XX		XX						
21		XX	XX	XX	XX			XX			XX	
22				XX	XX			XX	XX	XX	XX	
23	XX				XX				XX	XX	XX	XX
24	XX					XX	XX		XX	XX		XX
25	XX		XX			XX	XX					XX
26		XX	XX	XX		XX	XX	XX				
27		XX	XX	XX	XX			XX			XX	
28				XX	XX			XX	XX	XX	XX	
29	XX				XX				XX	XX	XX	XX
30	XX					XX	XX		XX	XX		XX

Note. "XX" indicates that the class was tested on the week shown.

The number of exercises tested each week varied from 8 to 15. Three forms of the test were given each week, each covering the same objectives. On each form a given objective was represented by a different exercise fitting the description of the objective. Each test included a varied sample of the entire range of the second-grade objectives, which include various counting questions, addition and subtraction in horizontal, vertical, oral, and word problem format, and a few elementary exercises involving multiplication, division, measurement, and time.

In this study only those thirteen objectives in the category 'vertical addition' are used. These objectives are assessed a total of 50 different times during the 26 weeks of testing. Further description of the exercises is given in Section IV-2.

III-3. A performance model for addition

The model developed in this study and tested with the weekly test data from Nicaragua has two parts. One is a performance model based upon an analysis of the structural components of vertical addition exercises. The other part of the model involves the application of mathematical learning theory to the performance model. We begin with the performance model.

Let us assume that correctly answering an elementary addition exercise consists of performing certain independent operations successfully. Then the probability that a given student will correctly answer a given exercise is the product of the probabilities of his correctly performing each individual operation. The analysis presented here considers three operations - adding two digits together, moving to the next column, and carrying. Thus any exercise may be analyzed for the number of times each of these operations is required, and a prediction made of the probability of answering it correctly.

For example, consider the following vertical addition exercise:

$$\begin{array}{r} 431 \\ +\ 297 \\ \hline \end{array}$$

There are <u>three</u> two digit additions required (excluding the addition associated with the carry), <u>two</u> moves to a new column, and <u>one</u> carry. Thus a prediction for the proportion correct (P_n) on this exercise, when taken on the nth week is :

$$P_n = P_{1n}^{3} P_{2n}^{2} P_{3n} \qquad \text{(Equation 3)}$$

where P_{in} is the probability of correctly performing the i_{th} operation on week n.

It must be noted that the assumption of independence of the individual sub-operations is a very critical part of the model because it greatly simplifies the above expression for predicted performance correct. If conditional probabilities were to be employed in the expression for P_n, the analysis as outlined here would become impossible because of the introduction of new parameters. However it is appropriate in an initial attempt such as this in the development and study of a model to seek out simplifying assumptions which are intuitively satisfying. The assumption of independence, like the assumption of continual reinforcement, seems intuitively defensible, although any proof or disproof would certainly require extensive study.

III-4. Introducing learning into the model

Since instruction, practice, and assessment on addition objectives occur regularly throughout the year in this curriculum, it is natural to introduce elementary learning theory into the performance model. If we assume that instruction and practice represent reinforcement in a given operation, then it is reasonable to express the increase in ability (i.e., increase in probability of performing an operation correctly) according to the following relationship, which is a slight modification of the first version of equation 1 (or equation 2a) of Section II:

Let p_n = probability of correctly performing a given operation on week n

and Θ = learning parameter ($0 < \Theta <= 1$)

then $p_{n+1} = p_n + \Theta(1 - p_n)$ (Equation 4)

Thus each week the ability to perform each operation is assumed to increase by an amount proportional to the difference between the current value of the ability, and its maximum value, in this case 1.

It can be shown that given the linear assumption of equation 4, there exists an explicit expression for p_n in terms of initial conditions and the week n. The result, which can easily be proved by mathematical induction, is:

$$p_n = 1 - (1 - p_1)(1 - \Theta)^{n-1} \quad \text{(Equation 5)}$$

where p_1 = the initial level of the ability.

This expression has been used in studies of latency, or number of trials before a correct response, and in applications of Bush and Mosteller's linear operator model to repeated trials with reinforcement (Coombs, Dawes, and Tversky, 1970; p. 261 ff). If the latency study operates under the assumption that a trial consists of all the responses up to and including the first correct (reinforced) one, there is indeed reinforcement on every trial, and the direct expression for p_{n+1} may be applied instead of the recursive one. The expression could be be applied to a simple T-maze experiment, for example, if a trial was defined to be one on which a response was reinforced, with other choices ignored. Then equation 5 could be applied to that situation.

The assumption of independence of the three sub-operations brings us to the final part of the formulation of a predictions of proportion correct. Let us return to the particular addition exercise presented in the previous section. Equation 3 gives the model's prediction of probability correct recursively in terms of the current levels of ability on the three sub-operations. Since we make no distinction among individuals, this equation also gives the predicted proportion correct for the group. Applying to this the explicit expression of equation 5 we have the following prediction of the proportion correct for this particular exercise on week n:

$$P_n = (1 - q_{11}\, a^{n-1})^3\, (1 - q_{21}\, a^{n-1})^2\, (1 - q_{31}\, a^{n-1}) \qquad \text{(Equation 6)}$$

where $1 - q_{i1} = p_{i1}$ = the initial level of ability on the i^{th} sub-operation

and $a = 1 - \theta$

The learning parameter θ is assumed in the model to be constant for an individual, but variable throughout a population. The available data, however, consist only of individual responses to items which are composed of the three individual component operations. In classical experiments with simple learning the parameter could be estimated from the form of the performance vs. time learning curve. Since each item contains a different combination of the three sub-operations, it is not possible to estimate an individual's change in ability to perform the sub-operations, and therefore it is not possible to estimate a learning parameter for each student.

Thus the basic analysis will be quite different from that normally encountered in mathematical learning theory. The average value of the parameter θ will be estimated for the entire sample of students from the proportion correct and the structure of the exercises on a given week. A variation of this approach is to calculate a learning parameter for each class and average across classes. This modification will be discussed in detail in a later section.

III-5. Estimating the parameters

A. The initial skill levels

Of the four parameters of the model, three are the initial probabilities of performing an operation correctly, and one is the learning parameter, θ . The three initial skill levels may be estimated from observed proportions correct during the early weeks of the year. This estimation process requires a set of three exercises which involve all three of the sub-operations in varying degree. We write equation 6 with n=1 and P_n equal to the three respective observed proportions correct, and solve for p_{11}, p_{21}, and p_{31}. That is, if b_i , m_i, and c_i represent the numbers respectively of two digit additions, moves to a new column, and carries on exercise i, and P_i is the observed proportion correct on exercise i , then the following equation may be written for each exercise i in a given set of exercises:

$$P_i = (1 - q_{11})^{b_i} (1 - q_{21})^{m_i} (1 - q_{31})^{c_i} \qquad \text{(Equation 7)}$$

where $1 - q_{j1} = p_{j1}$ = initial skill level on the j^{th} sub-operation

Not any set of three exercises will be suitable for this estimation, for it is possible that a set of equations will not yield a unique solution for three variables. For example, if the values of b_i, m_i, and c_i on one exercise are a linear combination of the values on another, then

study were being designed from the beginning, it would be important to ensure that a suitable set of exercises is available early in the year to make this estimation possible. In the present study, however, we must look for an appropriate set of exercises and let that week become the first.

B. The learning parameter

One parameter remains to be estimated - the learning parameter θ. This is accomplished here by means of the method of maximum likelihood invented by R.A. Fisher in 1912. In this method, a parameter θ of the distribution of a random variable is estimated from independent observations $\{X_i\}$ of the variable. The likelihood, $L(\{X_i\} \mid \theta)$, is defined as the joint probability density of the set of observed values $\{X_i\}$ conditioned on θ. That is, $L(\{X_i\} \mid \theta)$ is the probability of obtaining this particular set of observations $\{X_i\}$ for a given value of θ. We seek the value of the parameter θ which maximizes this probability.

This method of parameter estimation is employed by Suppes and Atkinson (1960) in studies of mutually adaptive two person interactions, and by Suppes (1973) and Suppes and Morningstar (1972) in studies of elementary arithmetic exercises (without learning).

The procedure for obtaining the maximum likelihood estimate of parameter θ for any given exercise proceeds in a straightforward manner, and the result is appealing to the intuition: the estimate of θ is the value which makes the predicted proportion correct equal to the observed proportion correct. That is, if we consider again the sample item presented in section III-3, the maximum likelihood estimate of the parameter θ for this exercise week n is the solution of the equation 6, a <u>prediction</u> equation, with P_n set equal to the <u>observed</u> proportion correct. Further discussion of this result

is presented in the next section. The proof proceeds as follows:

If P is the predicted proportion correct of equation 6 (and the probability of a correct response from an individual), and N/M is the observed proportion correct (for this derivation, N is the number of students responding correctly to the item, out of a total of M students), then

$$L(\{X_i\} \mid \theta) = C(M,N)\, P^N\, (1-P)^{M-N}$$

where C(M,N) = the number of subsets of N elements which can be chosen from a set of M elements

= the number of different combinations of responses which would produce the observed proportion correct, N/M

To maximize L over θ, we set the derivative equal to zero:

$$\frac{\partial L}{\partial \theta} = C(M,N)\left[N \cdot P^{N-1}(1-P)^{M-N}\frac{\partial P}{\partial \theta} + (M-N)(1-P)^{M-N-1}P^N\left(-\frac{\partial P}{\partial \theta}\right)\right.$$

If $\frac{\partial P}{\partial \theta} \neq 0$, then $\frac{\partial L}{\partial \theta} = 0$ when:

$$N P^{N-1}(1-P)^{M-N} = (M-N)(1-P)^{M-N-1} P^N$$

From this it follows that:

$$\frac{1-P}{P} = \frac{M-N}{N} = \frac{1-\frac{N}{M}}{N/M} \quad \text{OR} \quad \underline{\underline{P = \frac{N}{M}}}$$

III-6. Judging the model

The primary means of judging the applicability of the model to this data is to study how well it predicts actual observed performance on a given item on a given week. Two different kinds of predictions are proposed:

- "Next week predictions" - here predictions of proportion correct for all the items on week n are made (from an equation like 6 of section III-4 using the average of the weekly estimates of the learning parameter up to week n-1. We can study how the accuracy of these predictions varies over time, how good they are on the average, and how the accuracy depends upon the type of exercise.

- "Far future predictions" - here a prediction of performance on a particular exercise on a given week n may be made from observations up to any other week j, where $j < n$. We may observe then how far into the future the model can be used to make reasonable predictions.

Another way of judging the model with this data is to study the stability of the weekly estimates of the learning parameter θ, which the model assumes to be constant over time.

'Next Week Prediction' also provides the means to assess another aspect of the model - the initial skill levels estimated for each of

the three component operations. That is, it is possible to vary these initial levels and observe the effect produced in the overall accuracy of the predictions.

In addition to observing the accuracy of these predictions, and studying the parameter estimation, the accuracy of the "next week" predictions can be compared with a regression approach similar to that described in Section II-2. Suppes, Hyman, and Jerman (1967) report remarkably accurate fit in predicting performance on exercises, given a few structural features of the exercise. Many structural features were considered, and those few which gave the best fit were retained. For example, for vertical addition, the three features used were the total number of steps required in the solution, and the magnitudes of the smallest and largest addends. In those analyses, proportion correct was the dependent variable and the structural features the independent variables.

The limitations of this approach have been discussed earlier, and in Suppes (1973). However, it does represent the best competing model to the learning theory/structural analysis model presented here because it is concerned with similar types of exercises and does incorporate structure of the exercises into the analysis.

The regression analysis thus could be a base line according to which the learning theory model could be evaluated. The former does not consider learning, but because the predictions it makes are often good, it provides a useful standard by which to judge the predictions

made by the learning model.

The regression predictions will be obtained from the following equation

$$p = a_1 b + a_2 m + a_3 c + a_4$$

where p is the proportion correct and b, m, and c are the numbers respectively of adding two digits, moves to new columns, and carries required by the exercise. The coefficients a_1, a_2, etc are obtained from a regression with observed proportion correct as the dependent variable and the structural features of the exercises as the independent variables. (Note: the variable p is defined to assume values from 0 to 1, but regression estimates, of course, could fall outside this range. Thus it is customary in work of this kind to employ a probability preserving transformation on the proportion correct, do the regression, and then apply the inverse of the transformation to the regression estimates. The transformation used here is $\log((1-p)/p)$.

III-7. Summary of the Analysis of the Model

A. The goal of this analysis is to test an original model for performance of elementary addition exercises. The model combines learning theory with a structural analysis of exercises.

B. The data consist of proportion correct on 13 structurally different vertical addition exercise types. The testing took place in 12 different second-grade classrooms during 30 consecutive weeks of the 1976 school year, as part of the Radio Mathematics Program in Nicaragua. The exercises considered here were all presented on individual student papers and account for from one to four items of the 10-15 items given each week.

C. The assumptions of the basic model are as follows:

- The probability that a given individual will answer a given exercise correctly may be predicted from his/her ability level on three separate sub-operations involved in the exercise.

- The proportion of students in the sample who correctly answer the exercise is predicted by the individual probability.

- The individual sub-operations are independent in the probability sense: i.e., the probability of correctly performing a series of them, including a single one more than once, is equal to the product of the respective probabilities.

- The level of ability on any sub-operation increases with reinforcement according to a linear learning model.

- This particular curriculum allows the assumption that reinforcement is continual.

- Each level of ability increases with the same (constant) learning parameter.

D. Predicting proportion correct from the model:

- Each exercise type is characterized in terms of the number of each of three sub-operations which are required.

- Assuming independence of the three sub-operations, a prediction on any week n may be made for the proportion correct that will be obtained on each exercise. There are four parameters in this prediction: the skill level on week n for each sub-operation, and the learning parameter, theta, whose magnitude determines how much the three skill levels improve each week.

- According to linear learning theory, the level of ability on on each sub-operation on week n may be expressed as a linear transformation of the ability on the previous week (n-1). An expression for the ability on week n may be derived which gives this value directly in terms of the initial level of the skill and the week, n. The result is that a prediction of

the three initial skill levels, the learning parameter theta, and n. (An example of such a prediction is given in section III-4, equation 6).

E. Estimation of the parameters

- The three initial skill levels are estimated from performance on exercises given in the early weeks of the year.

- An estimate of the learning parameter is made each week using the method of maximum likelihood. This is possible because it can be shown that the value of the learning parameter which makes the predicted outcome on an exercise equal to the observed outcome is the maximum likelihood estimate of the parameter.

- The actual estimation of the learning parameter is made by setting an expression such as the right side of equation 6 of section III-4 (the prediction made by the model of proportion correct for a given exercise on a given week) equal to the actually observed proportion correct. The result is a polynomial in θ which can be solved by Newton's Method.

F. Judging the Model

The model will be judged by how well it makes both short and long range predictions of proportion correct on a given exercise, and by the stability of the estimates of the learning parameter θ.

In addition, the predictions of proportion correct will be compared to those of a regression of proportion correct with structural features of the exercises.

III-8. Modifications of the model

Each of the following modifications to the basic model can provide further insight into the suitability of this kind of approach to test data collected over time. The first proposed modification is the most obvious one: we shall study the change in accuracy of the predictions as the number of individual operations is reduced by eliminating one or two of them from the analysis.

Second, the cumbersome process of estimating the learning parameter may be replaced by using a constant value for it throughout the year. Of course the linear learning model assumes that this parameter is constant, but in this analysis its estimate is updated with each new week's observations. However, after the analysis has been carried out, we may repeat the entire prediction process using the average of the estimates of the learning parameter to estimate the increase in ability each week. Another modification in the analysis is described below.

A. Analysis by class

Recently Cronbach (1976) has raised the question of the appropriate unit of analysis in educational research. He cites instances in which, for example, significant aptitude-treatment interactions were obtained when the analysis grouped all students together, but disappeared when class membership was taken into account. He makes two points which are particularly important to the

present study.

First, ignoring class membership implies that students within a class are independent of one another, and not influenced by the class (whether indirectly, as by geographical location, or directly, as by class atmosphere). Second, he urges that the "unit of analysis" should correspond to the "unit of treatment". The curriculum of the Radio Mathematics Program is even more obviously a classroom-oriented presentation than the examples which Cronbach gives: all students in each room are exposed to exactly the same instruction; any distraction, or discipline problem will likely affect all individuals in the class; an individual teacher's attitude toward the radio lessons will affect the entire class in roughly the same way; and there is no individualization in the radio lessons.

Thus classes may have identifiable traits which might cause the individual members to respond differently across classes, and any analysis ignoring class membership will lose this effect. The basic analysis described here treats grand means of all the individuals throughout 12 classes. An alternative would be to predict for each class the outcomes on the exercises given to its members. The analysis would then be able to take into account any differences in either initial ability or learning rate which exist across classes.

IV. EXPLICIT FORMULATION AND RESULTS

IV-1. Notation

Parameters of the Model:

A. Skill parameters (probabilities associated with skill levels on sub-operations)

$1 - \beta$ = initial probability of performing an addition of two digits (e.g. 2+5 or 8+6) correctly

$1 - \mu$ = initial probability of moving successfully from one column to the next

$1 - \gamma$ = initial probability of carrying correctly (no distinction is made between carrying 1 and carrying 2, nor is the act of adding the carried number considered a new addition)

B. Learning parameter

θ = the learning parameter of equation 4. Each of the above probabilities are assumed to increase each week according to equation 4.

$\alpha = 1 - \theta$

Exercise Characteristics:

b = Number of additions required by the exercise

m = Number of moves to the next column required by the exercise. The number of columns in a vertical addition exercise is the same as the number of digits in the largest addend.

c = Number of carries required

Other Variables:

N = number of students responding to an exercise

n = number of students correctly answering an exercise

P = n/N = proportion correct on an exercise

j = number of weeks since the week in which the initial skill parameters were calculated (week 9 in the present study)

Thus the prediction of the proportion correct for a given exercise, in this notation, is:

$$P = (1 - \alpha^{j}\beta)^{b} (1 - \alpha^{j}\mu)^{m} (1 - \alpha^{j}\gamma)^{c}$$

(equation 8)

where $\alpha = 1 - \theta$

IV-2. Analysis of exercises

Although addition was taught and tested in various formats in the second grade curriculum - vertical, horizontal, oral, and word problem - the data used in this study is for the 13 vertical addition objectives which were tested in 1976.

Each objective was tested by three different but comparable exercises, each given to approximately the same number of students. The proportion correct for each objective is taken in this study to be the average of the student scores on the three exercises, weighted according to the number of students receiving each form. The three different exercises were distributed among the students in each class, so any statistical interaction between exercise characteristics and class characteristics may be ignored. Giving three different forms in a single class also reduces the effects of students copying answers from one another.

"Vertical" refers to the appearance of the exercise on the student test paper, examples of which are included in Appendix A. The sample student test papers, which are the three forms for week 15, show the manner in which vertical addition exercises were presented. The tests contained from 1 to 4 vertical addition exercises, except that on three weeks there were none.

All three forms of each week's test were identical in format. Most of the vertical addition exercises were presented more than once

during the year, and in these cases different but comparable exercises were involved. Throughout the year a total of 50 vertical addition exercises were presented. Table 2 contains a list and description of the objectives.

The three columns headed b, m, and c in Table 2 contain the analysis of the operations required for each exercise, as described in part 1. of this section. The following criteria were followed in the assignment of b, m, and c:

b = number of times an addition is carried out, not including the addition of 1 or 2 involved in carrying.

Thus for AV04, for which a typical exercise would be 234 + 163 (presented in vertical format), b = 3 since a total of three separate additions are required. In general if a problem contains x addends with y columns each, then b = y*(x-1). Objectives AV03 and AV10 present a different situation because the addends do not have the same number of digits. We take the value of b to be 2 for an exercise such as 23 + 5 because there is an implied addition of zero, as if the exercise were 23 + 05. This is particularly appropriate in this curriculum, since in Nicaragua leading zeroes are retained in various calculations.

m = number of moves to a new column. The value of m has been taken in this study to be one less than the number of digits in the largest addend.

Some exercises involve 'passing', the term used in this

Table 2

Description of Vertical Addition Objectives

Objective	Description	No. of additions (b)	No. of moves to a new column (m)	No. of carries (c)
AV01	A+B<=9	1	0	0
AV02	AB+CD, without passing or carrying	2	1	0
AV03	AB+C, A+BC, without passing or carrying	2	1	0
AV04	ABC+DEF, without passing or carrying	3	2	0
AV09	AB+CD, carrying, but not passing	2	1	1
AV10	AB+C, carrying, but not passing	2	1	1
AV12	ABC+DEF, carrying two times, but not passing	3	2	2
AV19	A+B>=10	1	1	0
AV20	AB+CD, passing	2	2	0
AV23	AB+CD, carrying and passing	2	2	1
AV27	AB+CD+EF, AB+CD+E, carrying 1	4	1	1
AV29	AB+CD+EF, AB+CD+E, carrying 1 or 2	4	1	1
AV30	ABCD+EFGH+IJKL, carrying 1 or 2	8	3	3

Note. "Passing" occurs when the answer contains more digits (columns) than any of the addends. Example:

```
  35
+ 82
-----
```

curriculum for the need to enter a digit in a column to the left of those in the addends. AV20 is one of these (example: 67 + 52), since the answer contains a digit in the hundreds place, but the addends are less than 100. For AV20 we set b = 2 and m = 1.

c = number of carries.

Here the size of the carry (1 or 2) is not distinguished. Also note that the operation of carrying is understood here to include the addition of the carried digit. We could have incremented b by 1 for every carry, but have chosen not to do so, in order to avoid creating a dependence between the value of b and the value of c. Also, passing is not considered to be a carry.

Note that for this set of exercises, b and m are nearly always dependent. That is, b = m + 1 except for exercises AV27, AV29, and AV30, which involve more than two addends. Since these exercises are tested only 6 times all together, this dependence may mean that one of the parameters may be eliminated in the analysis with no loss in predictive power. This is exactly what happens, as shall be seen, but it must be recalled at that time that the particular set of exercises included in this study may have produced this reduction of parameters.

Table 3

Vertical Addition Exercises Included

in First 5 Weeks of Testing

Week	Objective	No. of additions (b)	No. of moves to a new column (m)	No. of carries (c)	No. of students tested (N)	Porportion correct over N students (p_i)
5	AV02	2	1	0	64	.95
6	AV04	3	2	0	81	.77
7	AV03	2	1	0	117	.90
7	AV02	2	1	0	117	.92
8	AV04	3	2	0	116	.79
9	AV03	2	1	0	99	.89
9	AV09	2	1	1	99	.33
9	AV27	4	1	1	99	.26

IV-3. Estimating the Parameters

A. The initial skill levels

As discussed in Section III-5, the initial skill levels are estimated from observations during an early week of the year. Table 3 contains the list of vertical addition exercises which were included in the first 5 weeks of testing, weeks 5 through 9.

The set of exercises presented on week 9 are suitable for estimating the the three initial skill levels. The procedure is straightforward: the set of estimates is the solution set of three equations, one version of equation 8 for each exercise on week 9. The results are:

$$\beta = .1$$
$$\mu = 0$$
$$\gamma = .6$$

Recalling that these parameters are the probabilities of making an error on the respective operations, we have these estimates of the initial skill levels, or probabilities of performing the operations correctly:

$1 - \beta = .9$ = Probability of doing a single addition correctly

$1 - \mu = 1.0$ = Probability of moving successfully from one column to the next.

$1 - \gamma = .4$ = Probability of correctly performing a carry

B. The learning parameter

As discussed in Section III-5, a maximum likelihood estimate of θ may be obtained by substituting the observed proportion correct on an exercise, n/N , for the predicted proportion, P, in equation 8. That is, the following equation must be solved for α :

$$n/N = (1 - \alpha^j \beta)^b (1 - \alpha^j \mu)^m (1 - \alpha^j \gamma)^c$$

$$(\alpha = 1 - \theta) \qquad \text{(Equation 9)}$$

A proof that the solution of this equation is the maximum likelihood estimate of the learning parameter is given on page 38.

In estimating the parameter θ for each exercise in the present model, Newton's Method was used to solve equation 9, to three place accuracy, for the jth power of $\alpha = 1 - \theta$. The next step is to take the jth root of that estimate as the estimate of $1 - \theta$. The initial value used for the root was 1.0, and a three place estimate of the jth power of α was obtained i under five iterations in most cases, with a high of 13 iterations required just once. Table 4 contains a summary of the results of this estimation process.

Figure 2 contains more information about these estimates of θ. The plot of θ vs. week shows that 35 of the 42 values

Table 4

Estimating for Each Exercise

Week	Objective	b	m	c	P_i	α^{j} *	K**	θ^{+}	Average θ
10	AV01	1	0	0	.90	1.000	1	.000	
11	AV19	1	1	0	.91	.900	2	.052	
11	AV02	2	1	0	.91	.461	3	.322	
11	AV12	3	2	2	.54	.371	5	.392	
									.255
12	AV09	2	1	1	.56	.608	3	.153	
12	AV04	3	2	0	.83	.602	3	.156	
12	AV23	2	2	1	.52	.671	3	.125	
									.145
13	AV03	2	1	0	.88	.619	3	.113	
13	AV19	1	1	0	.85	1.500	2	-.106	
									.003
14	AV09	2	1	1	.52	.671	3	.077	
14	AV20	2	2	0	.73	1.456	3	-.078	
									-.000
15	AV19	1	1	0	.83	1.700	2	-.092	
15	AV01	1	0	0	.83	1.700	2	-.092	
15	AV12	3	2	2	.48	.433	5	.131	
15	AV27	4	1	1	.57	.500	4	.110	
									.014
17	AV10	2	1	1	.52	.671	3	.049	
17	AV23	2	2	1	.53	.655	3	.052	
									.050
18	AV01	1	0	0	.95	.500	2	.075	
18	AV09	2	1	1	.72	.372	4	.105	
18	AV27	4	1	1	.69	.344	4	.112	
									.097
19	AV19	1	1	0	.88	1.200	2	-.018	
19	AV10	2	1	1	.77	.302	4	.113	
19	AV02	2	1	0	.94	.305	3	.113	
19	AV12	3	2	2	.68	.241	5	.133	
									.085

(continued on next page)

(Table 4, cont.)

Week	Objective	b	m	c	p_1	α^{j*}	K**	θ^+	Average θ
20	AV27	4	1	1	.80	.213	4	.132	
20	AV04	3	2	0	.87	.453	3	.070	
									.101
21	AV10	2	1	1	.66	.458	3	.064	
21	AV23	2	2	1	.67	.443	3	.066	
21	AV12	3	2	2	.71	.215	5	.121	
									.084
22	AV03	2	1	0	.80	1.056	2	-.004	
22	AV27	4	1	1	.71	.319	4	.085	
23	AV12	3	2	2	.58	.331	5	.076	
23	AV23	2	2	1	.70	.400	3	.064	
									.070
24	AV29	4	1	1	.61	.446	4	.053	
25	AV10	2	1	1	.65	.472	3	.046	
25	AV23	2	2	1	.63	.502	3	.043	
25	AV30	8	3	3	.67	.149	10	.113	
									.067
26	AV01	1	0	0	.86	1.400	2	-.019	
26	AV27	4	1	1	.78	.236	4	.082	
									.031
27	AV12	3	2	2	.70	.224	5	.080	
29	AV20	2	2	0	.89	.566	3	.029	
29	AV23	2	2	1	.83	.220	4	.073	

*j = Week less 9.

**K = Number of iterations required in Newton's method (estimate Jth power of α by Newton's method on equation 8, 3 place accuracy).

+ θ = 1 - jth root of above estimate.

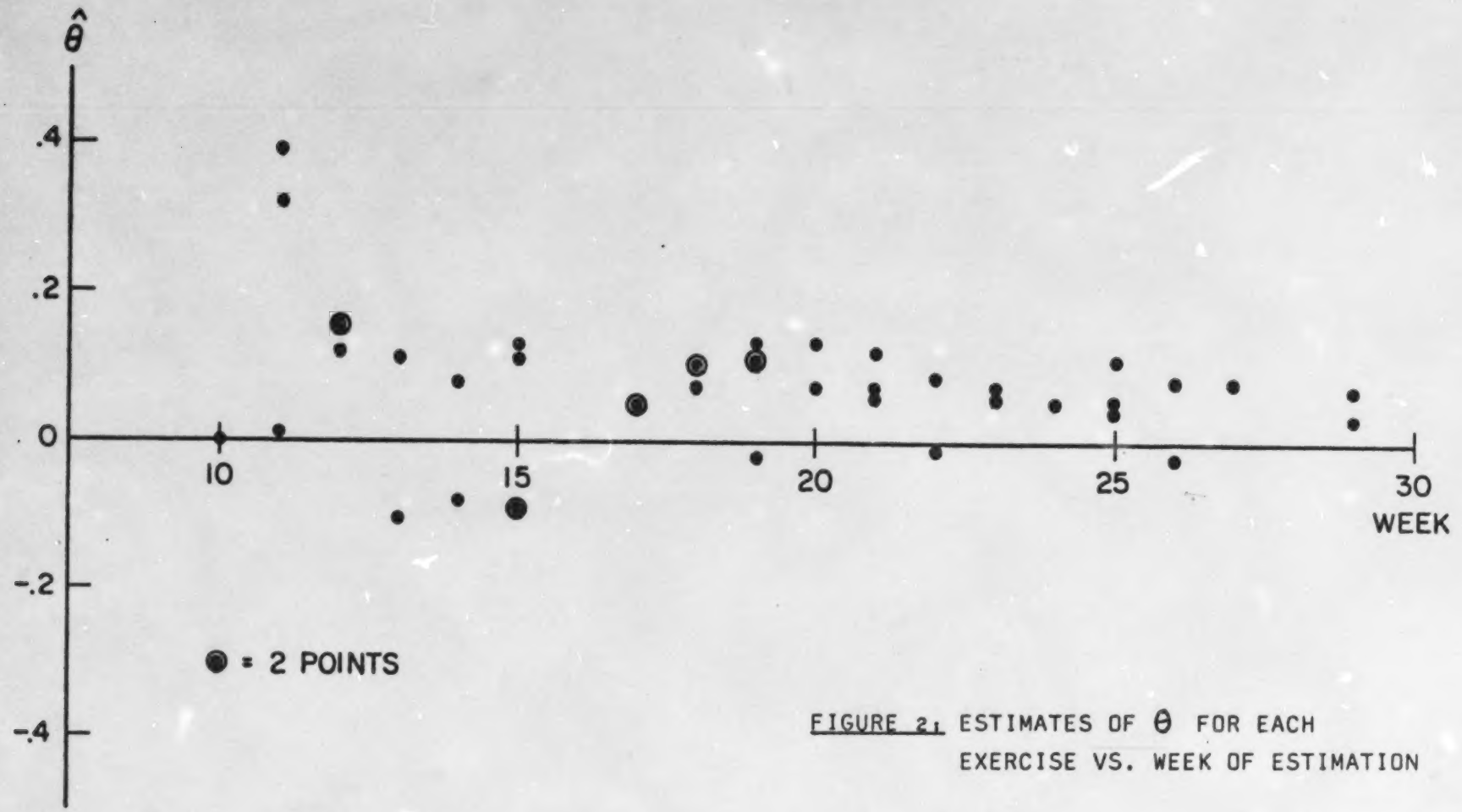


FIGURE 2. ESTIMATES OF θ FOR EACH EXERCISE VS. WEEK OF ESTIMATION

of θ (83 %) lie between -0.02 and 0.16. Moreover, all of the estimates after week 15 are within this interval. Table 5 presents a comparison of the results for all 18 weeks, and the results for the last 12 weeks.

Negative estimates of θ are unexpected occurrences; equation 9 will produce a negative value of θ ($\alpha > 1$) if the observed proportion correct is smaller than the prediction from equation 8 with n = 0. That is, on these exercises performance is worse than would be expected at the beginning of the year. This occurs on 6 out of 42 exercises.

In a summary of stochastic learning theory studies with animal and simple task learning, Coombs, Dawes, and Tversky (1970, p.264) report that estimates of θ usually are below 0.2, which is consistent with these results.

Although the stability of the estimates of θ in the later weeks may be an indication that the model does better then, there is another explanation. Recall that the estimate of θ involves taking the jth root of the Newton's Method approximation. This will tend to reduce the range of values as j gets large. For example, if x ranged from .4 to 1.2, then the 15th root of x will range between .94 and 1.01. Thus we must wait for further evidence of the quality of the model. In part 8 we will study the model by attempting to predict exercise outcomes.

These estimates of the learning parameter θ are made

Table 5

Estimates of θ for Each Exercise

Period studied	Range	Mean	Standard deviation
All 18 weeks (10 to 29)	(-0.11, 0.39)	0.076	0.09
Last 12 weeks (17 to 29)	(-0.02, 0.13)	0.071	0.04

Note. There were no vertical addition exercises on weeks 16, 28, and 30.

for each exercise, but we really seek an estimate of θ for a given week, and need some way to combine the estimates of the parameter over all the exercises taken in a given week. This matter is taken up in the next section.

IV-4. Estimating θ For a Week

In the previous section estimates were obtained for the learning parameter θ from each vertical addition exercise given during the year. The basic analysis which was outlined in Chapter III, however, involves obtaining an estimate for θ for each week. Hence some means for combining the estimates across exercises on a given week is needed.

In Appendix B the maximum likelihood estimation process is carried out for an arbitrary number of exercises on a given week, and it is shown there that no tractable expression is available for θ across exercises. In principle it would be possible to use the expression derived there to produce an estimate of θ by Newton's Method, but the process for even two exercises in a week becomes very cumbersome. So we will take another approach.

Since each exercise on a given week is taken by the same number of students, it is natural to expect that the simple average of the estimates of θ for each exercise will be a good approximation of the maximum likelihood estimate of θ for that week. We may judge how close the mean is to the maximum likelihood estimate by tabulating the expression for the likelihood which is given in Appendix B, noting which value of θ produces the maximum value for the week, and comparing this value with the mean of the estimates of θ (recall that these are MLE estimates) across exercises.

An appropriate range of test values of θ has been obtained by first taking the values 0 to .9 in steps of .1 and ensuring that there is only one peak in the L vs. θ curve in that interval. These preliminary results are not presented here. The values of θ for which the likelihood was calculated are -0.1 to 0.5, in steps of 0.01. These values correspond to the range of estimates of θ which were actually obtained for each exercise (see table 5).

A maximum of the likelihood function was obtained within the specified interval each week. Table 6 gives the results for one of the weeks. The values of L are incredibly small because of elimination of the constant term C(N,n), the number of ways in which n correct responses could be obtained from a group of N responses.

The results for the remaining weeks are listed in table 7. In 14 out of 18 comparisons, the mean is within .02 of the value of θ which maximizes the likelihood function for a week. The average absolute deviation between the two is 0.02. Further, as we shall see in the next section, the predictions of proportion correct are little different when either of these weekly estimates of θ are used. Because of its greater simplicity in computation, therefore, the mean will be used in what follows.

Figure 3 contains some descriptive statistics of this estimate for each week. The range of values of the weekly estimates is 0.000 to 0.255, and after week 14 the range is 0.014 to 0.101. The mean of

all these weekly estimates is 0.067 and the standard deviation is 0.063.

Table 6

Values of Log (L) Vs. Θ

for Week 13

Log(L)	Θ
-103.489	.40
-102.066	.39
-100.681	.38
-99.334	.37
-98.024	.36
-96.750	.35
-95.513	.34
-94.312	.33
-93.148	.32
-92.018	.31
-90.925	.30
-89.867	.29
-88.844	.28
-87.857	.27
-86.905	.26
-85.988	.25
-85.107	.24
-84.262	.23
-83.452	.22
-82.678	.21
-81.939	.20
-81.237	.19
-80.572	.18
-79.943	.17
-79.351	.16
-78.797	.15
-78.280	.14
-77.802	.13
-77.362	.12
-76.961	.11
-76.600	.10
-76.278	.09
-75.997	.08
-75.758	.07
-75.560	.06
-75.405	.05
-75.293	.04
-75.224	.03
-75.201*	.02
-75.223	.01
-75.291	.00
-75.407	-.01
-75.571	-.02
-75.784	-.03
-76.047	-.04
-76.362	-.05
-76.729	-.06
-77.150	-.07
-77.626	-.08
-78.158	-.09
-78.748	-.10

Note. "*" indicates maximum value of Log (L).

Table 7

Comparison of θ which Maximized the Likelihood for a Week

(Observed M.L.E),

and

the Average of the MLE Estimates for a Given Week

(Predicted M.L.E)

Week	Number of exercises	Average (predicted)	Value which maximized L (observed)	Observed minus predicted
10	1	.00	.00	.00
11	3	.26	.36	.105
12	3	.15	.14	-.01
13	2	.00	.02	.02
14	2	.00	.05	.05
15	4	.01	.09	.08
17	2	.05	.05	.00
18	3	.10	.11	.01
19	4	.09	.11	.02
20	2	.10	.11	.01
21	3	.08	.09	.01
22	2	.04	.06	.02
23	2	.07	.07	.00
24	1	.05	.05	.00
25	3	.07	.07	.00
26	2	.03	.06	.03
27	1	.08	.08	.00
29	2	.05	.06	.01

Note: Average of absolute deviations (observed minus predicted) = 0.02 .

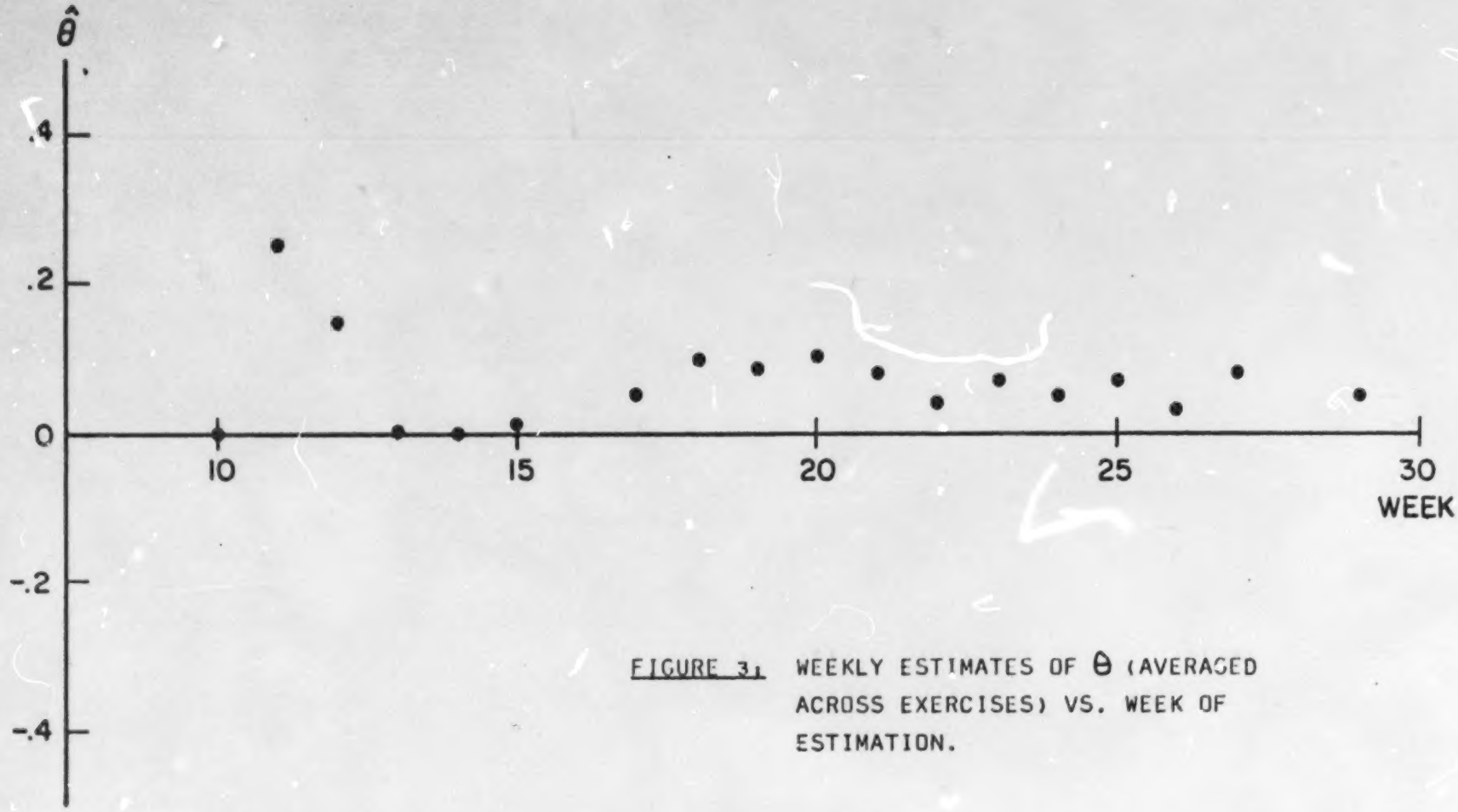


FIGURE 3: WEEKLY ESTIMATES OF θ (AVERAGED ACROSS EXERCISES) VS. WEEK OF ESTIMATION.

IV-5. Predictions

A model which can predict will be of more value than one which can merely explain. This model can be used to predict the proportion correct for exercises on future weeks in two ways: first, a prediction may be made for each exercise of each week using for the learning parameter, θ , the cumulative average of previous estimates. This cumulative average is the unweighted average of the weekly estimates.

Second, on each week the cumulative average of θ may be used to predict each of the exercises on all subsequent weeks. In other words, in the first method of prediction the proportion correct for an exercise on week j is predicted from equation 8 using the average of the estimates of θ for weeks 10 through $j-1$. This will be termed 'Next Week Prediction'.

In the second method of prediction, the results for an exercise on week j is predicted, again from equation 8, using the estimate of θ for week 10; it is then predicted again with the cumulative estimate of week 10 and 11, again from weeks 10,11, and 12, and so on. This will be referred to as 'Far Future Prediction', since on any given week predictions are made for exercises on all future weeks.

A. Next Week Prediction

Four measures are employed to assess the overall accuracy of

these predictions: the average of the absolute differences between observed and predicted, the correlation between the two, the average difference, and the proportion of the predictions which were greater than the corresponding observations.

The average of the absolute deviations is the most straightforward and useful of these measures; the correlation gives an idea of the consistency of the estimates, but of course the correlation can be high without the predictions being accurate. The last two measures are useful in detecting a bias in the prediction process.

The results for 'next week prediction' are as follows:

average absolute deviation = 0.09

correlation coefficient = 0.80

average deviation = -0.02

proportion of predictions > observed = 0.48

Thus on the average the prediction of the proportion correct for the group of students who took an exercise is highly correlated with the observed proportion and is within .09 of it. Also, the predictions are on the average unbiased. Recall that the weekly estimates of θ used here are averages of maximum likelihood estimates across exercises.

When the same predictions are made using instead the values of

θ which maximize the likelihood function for a week, the resulting measures of accuracy are 0.09, 0.76, 0.05, and 0.69, respectively. There is little difference between these sets of figures, although the second method has produces predictions which tend to be somewhat higher.

Figures 4 and 5 provide another way to look at the predictions. Figure 4 is a plot of predicted vs. observed values of proportion correct, and Figure 5 shows how the deviations between observed and predicted vary with time. The accuracy of the predictions is roughly the same throughout the year. The one outlier occurs on week 11, when a relatively complex exercise (b=3, m=2, c=2) has a prediction of 0.12, but an observed score of 0.54, for a difference of -0.42. In five subsequent predictions for this exercise type, the average absolute deviation is 0.12.

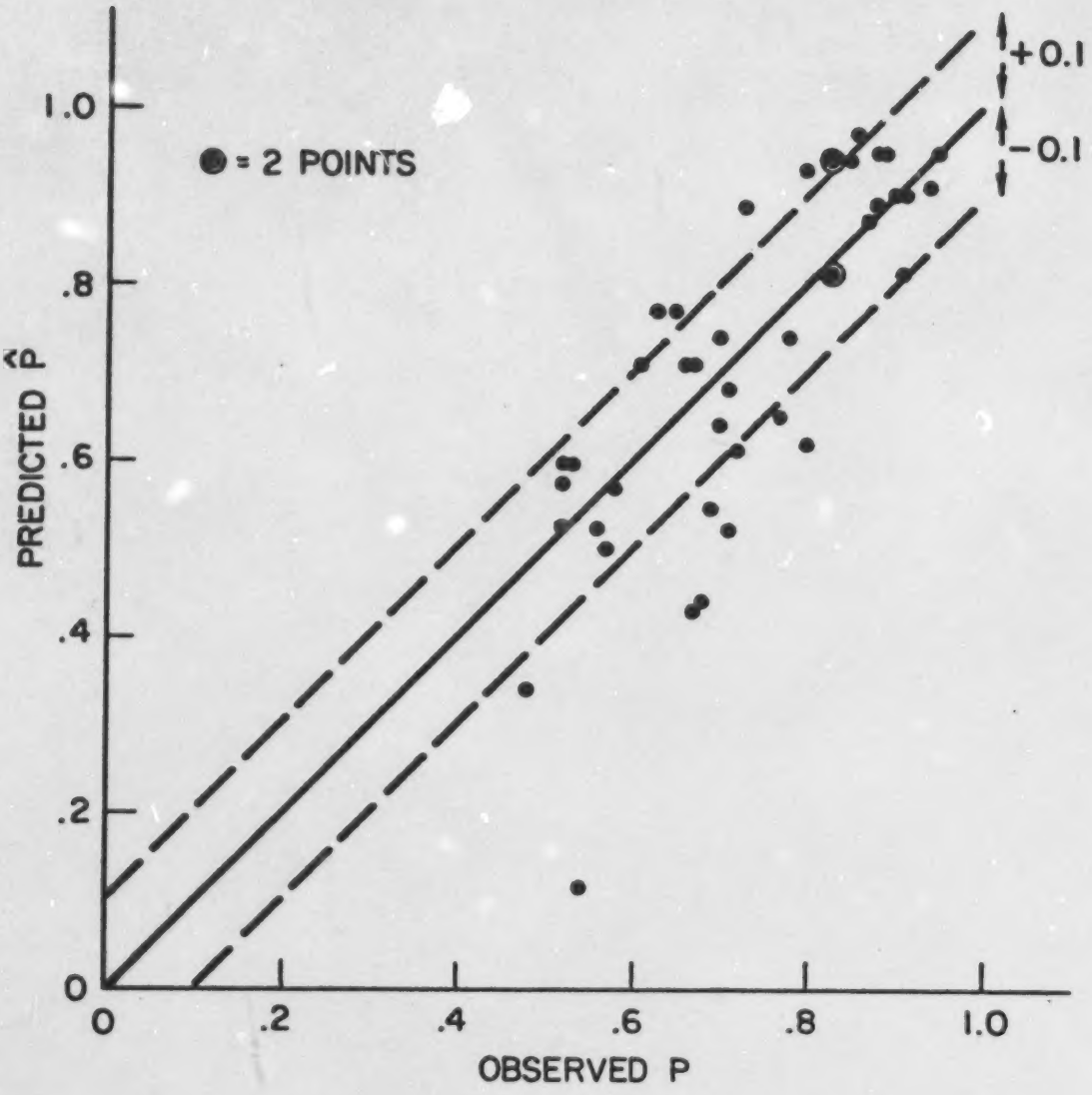


FIGURE 4: 'NEXT WEEK PREDICTIONS' VS. OBSERVED PROPORTION CORRECT

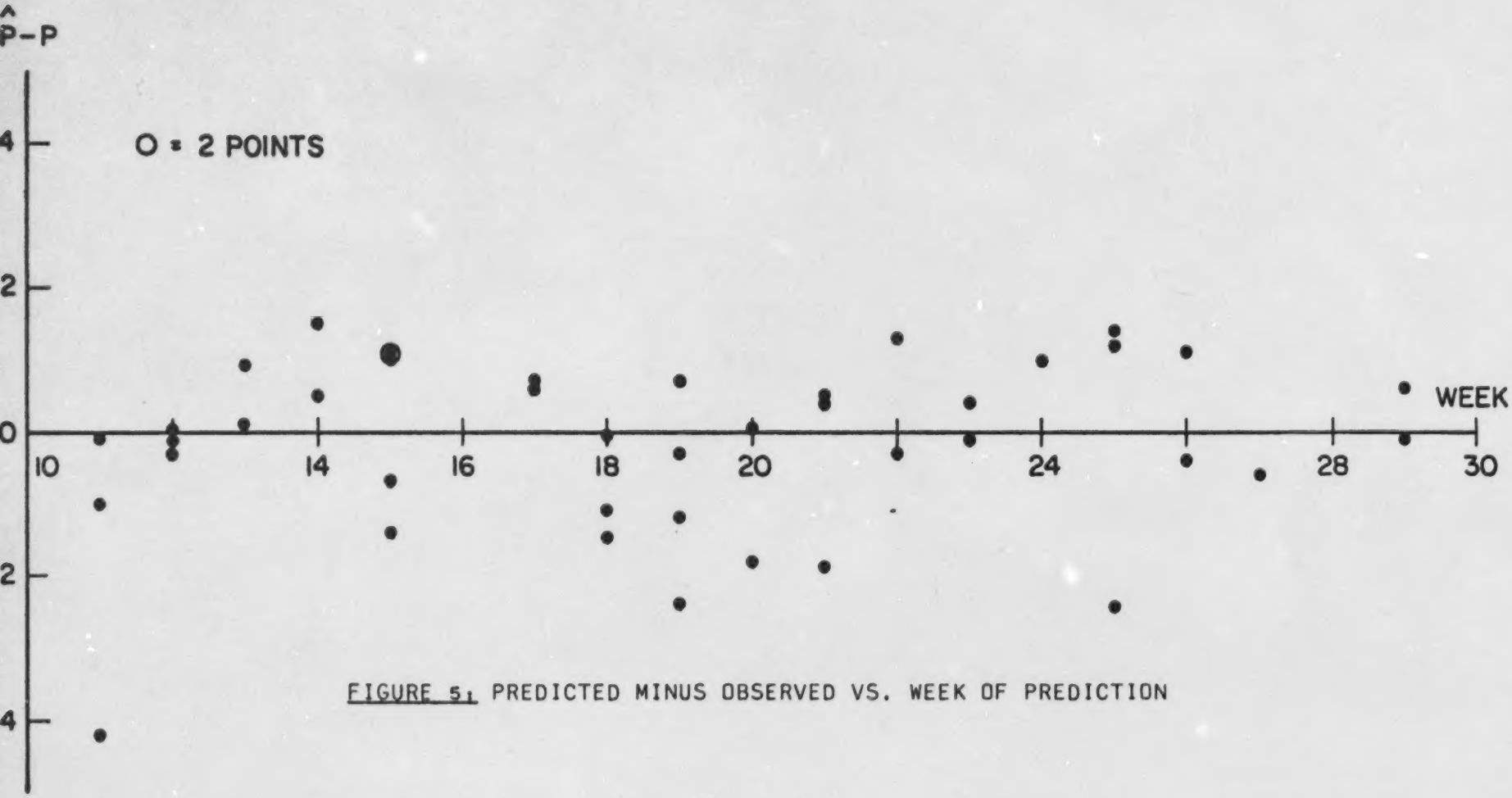


FIGURE 5: PREDICTED MINUS OBSERVED VS. WEEK OF PREDICTION

B. Far Future Prediction

Here a prediction of performance on a particular exercise on a given week j may be made from observations up to any other week j', where $j < j'$. We may observe then how far into the future the model can be used to make reasonable predictions.

The accuracy of these predictions is best illustrated by Figure 6, which shows the predictions made each week for selected exercises on weeks 25, 26, 27, and 29. The graphs show that in all cases the predictions are stable after about week 17. That is, the prediction for an exercise on week 29 made on week 27 is just about the same as the prediction based only on the data up to week 17; furthermore, both these predictions are off by less than 10 % . The stability of the estimates after week 17 is due to the stability of the estimates of θ , which is illustrated in figure 3 and table 5.

In Figure 6 the range of values of b, m, and c (the number of each of the three operations involved in each exercise) reveals that the particular exercises presented there form a good cross-section of the vertical addition exercises in this study. (Recall that m does not affect any prediction since the initial probability of doing its operation correctly was set to 1.0.) Note that exercise E is predicted least accurately. Its values of b, m, and c indicated that it is a unique type (see Table 4), and we would expect that such a rare exercise would be predicted with less accuracy.

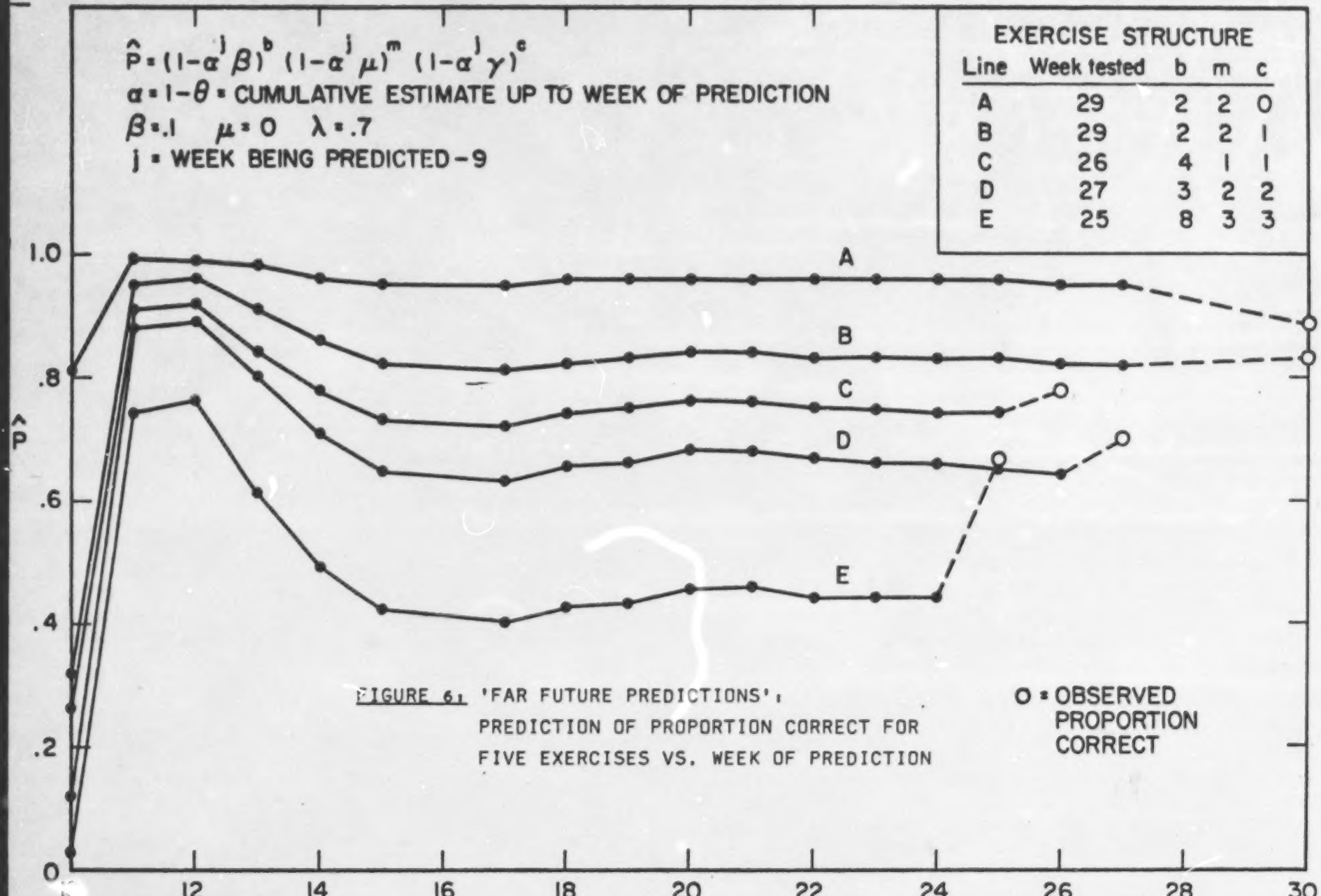


FIGURE 6: 'FAR FUTURE PREDICTIONS': PREDICTION OF PROPORTION CORRECT FOR FIVE EXERCISES VS. WEEK OF PREDICTION

C. Effect on the Predictions of Varying the Initial Skill Parameters

The focus of most of the preceding sections has been on the learning parameter θ. All of this discussion has been based on the values of the initial probabilities related to the three operations, which were derived in Section III-3. These values, which are the probabilities of not performing the operations correctly are:

$$\beta = .1$$
$$\mu = 0$$
$$\gamma = .6$$

The 'next week predictions' have been used to study these chosen values. The entire prediction process, beginning with the estimate of θ for each exercise, was carried out with different values of the skill parameters. The results, which are summarized in Table 8, indicate that the original choices for β and μ were the best possible ones, to the nearest tenth, although values of γ other than 0.6 do give similar results. Hence we can conclude that the original estimates of the skill levels were remarkably good, especially considering that they were obtained from three exercises on the 9th week of the year. Also, the accuracy of the predictions is much less sensitive to changes in γ than it is to changes in the other initial skill level parameters.

Table 8

Effects on 'Next Week Predictions' of Variations in Initial Skill Parameters, β, μ, and γ

β	μ	γ	r observed with predicted	Average absolute deviation	Average deviation	Proportion of prediction too high
.1*	.0*	.6*	.80	.09	-.02	.48
.2	.0	.6	.61	.18	.18	.95
.3	.0	.6	.33	.22	.22	.98
.1	.0	.6	.80	.09	-.02	.48
.1	.1	.6	.67	.10	.06	.76
.1	.2	.6	.52	.14	.09	.86
.1	.0	.1	.53	.11	-.04	.41
.1	.0	.2	.67	.10	-.02	.48
.1	.0	.3	.73	.09	-.02	.48
.1	.0	.4	.79	.09	-.01	.48
.1	.0	.5	.79	.09	-.02	.50
.1	.0	.6	.80	.09	-.02	.50
.1	.0	.7	.81	.09	-.02	.50
.1	.0	.8	.82	.10	-.03	.50
.1	.0	.9	.82	.10	-.03	.50

Note. "*" indicates values estimated from week 9.

IV-6 Predictions Using Modifications of the Model

A. Using fewer parameters

When fewer skill parameters are used, we are postulating that there there are fewer sources of error in solving an exercise. The predictions are again based upon equation 8, but if a parameter is eliminated, the value of its initial skill level (probability of making an error) is taken to be 0.

Table 9 shows the accuracy obtained as each parameter is eliminated in turn. The results are given for various values of the initial skill levels, although the theory does indicate which values of these should be used: the initial skill levels may be computed from the observations on week 9, as discussed earlier. (Of course this involves solving three equations for two or one variable - the estimates given are averages of the solutions obtained for each). The given are averages of the solutions obtained for each). The table indicates the results for the initial values estimated in that way.

As the earlier discussion indicated, the original model uses only two of these parameters, since μ was estimated as 0 on week 9, and this estimate was shown to be quite good in Section IV-5, where predictions are shown using various initial skill levels. Therefore when μ is eliminated the results marked with an asterisk in the table are those of the basic analysis.

Table 9

Results of Predictions Using Two Skill Parameters

Parameter eliminated: Gamma (carrying)					
Beta	Mu	Absolute deviation	Deviation	Proportion P > 0	r
.1	.0	.14	-.07	.36	.28
.1	.1	.11	.04	.64	.43
.1	.2	.13	.08	.76	.33
.2	.0	.17	.17	.91	.31
.2	.1	.19	.19	.93	.34
.2	.2	.20	.20	.93	.24
.3	.0	.22	.22	.98	.09
.3	.1	.22	.22	.98	.10
.3	.2	.23	.23	.98	.04
.3*	.3*	.23	.23	.95	.01

Parameter eliminated: Mu (moving to new column)					
Beta	Gamma	Absolute deviation	Deviation	Proportion P > 0	r
.1	.0	.14	-.07	.36	.28
.1	.2	.10	-.02	.48	.67
.1	.4	.09	-.01	.48	.77
.1*	.6*	.09	-.02	.48	.80
.2	.0	.17	.17	.91	.30
.2	.2	.18	.18	.93	.56
.2	.4	.18	.18	.95	.60
.2	.6	.18	.18	.95	.61

(Table 9, cont.)

Parameter eliminated: Beta (adding two digits)					
Beta	Gamma	Absolute deviation	Deviation	Proportion P > 0	r
.1	.0	.23	-.14	.26	.21
.1	.2	.12	-.06	.36	.67
.1	.4	.11	-.06	.41	.80
.1	.6	.12	-.07	.38	.84
.1*	.7*	.13	-.07	.38	.85
.2	.0	.12	.04	.60	.35
.2	.2	.10	.05	.67	.65
.2	.4	.10	.05	.74	.71
.2	.6	.10	.05	.76	.72
.3	.0	.14	.08	.74	.32
.3	.2	.13	.09	.81	.48
.3	.4	.13	.09	.88	.54
.3	.6	.14	.08	.91	.56

Note. "*" indicates values estimated from week 9.

The table shows that little accuracy is lost when two parameters are used, although the elimination of γ does reduce the correlation markedly, from 0.80 to 0.43 or less. This indicates that although the average error of the predictions is, on the average, not much different from that of the full model, the predictions do not tend to depend as much upon the observed values. This is because, without γ, the range of the predictions is markedly reduced.

Table 10 shows the prediction results when two parameters are eliminated. Again the estimates made from the observations on week 9 are indicated with an asterisk. The best results for the β model and the μ model are about the same in average absolute deviations (0.11 - 0.14) and correlation (0.25 - 0.28), and differ from the full model primarily in correlation. The λ model does remarkably well, however - its best absolute deviation (0.12 - 0.13) and correlation (0.81 - 0.86) compare quite favorably with the full model.

In summary, the parameter γ is most important in the analysis. A model of this kind which considers only carrying errors will predict almost as well as one which considers all three sources of error. This conclusion, however, may depend upon the particular grade level which is studied.

B. The model with a constant learning parameter.

The basic analysis uses the cumulative estimate of the learning

Table 10

Results of Predictions Using One Skill Parameter

Parameter Used: Beta (adding Two Digits)				
Beta	Absolute deviation	Deviation	Proportion P > 0	Correlation
.1	.14	-.07	.36	.28
.3	.22	.22	.98	.19
.5	.25	.25	.93	-.10

Parameter Used: Mu (Moving to a New Column)				
Mu	Absolute deviation	Deviation	Proportion P > 0	Correlation
.1	.11	.00	.43	.25
.2	.15	.09	.74	.19
.3	.18	.11	.79	.09
.4	.20	.11	.93	.05

Parameter Used: Gamma (Carrying)				
Gamma	Absolute deviation	Deviation	Proportion P > 0	Correlation
.1	.21	-.03	.57	.49
.3	.13	.01	.57	.73
.5	.12	-.01	.57	.81
.6	.13	-.02	.50	.83
.7	.13	-.02	.50	.85
.9	.15	-.04	.45	.86

parameter θ, whose estimation is computationally the most complex aspect of this analysis (see Sections III-3 and III-4). The learning theory upon which the model rests, however, assumes that the θ is constant. Therefore it is useful to observe the accuracy of the predictions under this restriction.

Table 11 shows that the overall accuracy of the predictions is not much changed with a constant θ in place of the average of the computed values. Also it can be seen that the correlation between the observed and predicted scores is relatively insensitive to changes in θ. The average absolute deviation, however, does vary between 0.09 and 0.27 as θ ranges from 0 to 0.10.

The predictions are consistently low for low values of θ, showing that the addition of the learning parameter to the model has contributed substantially to its overall accuracy.

C. Analysis by Individual Class

The procedure for predicting scores for individual classes was the following:

(1) The initial skill parameters, β, μ. and γ were estimated for each of the 12 classes in the manner described in Section IV-3, but based upon observations on weeks 11 and 12. (Week 9 could not be used as in the basic analysis because only half the classes were tested then.)

Table 11

Predictions Using Constant Learning Parameter

Theta	Correlation Observed/ Predicted	Average Absolute Deviation	Average Deviation	Proportion Predicted> Observed
.10	.80	.09	.04	.62
.09	.81	.09	.02	.62
.08	.82	.09	-.00	.57
.07*	.83*	.09*	-.03*	.45*
.06	.83	.10	-.06	.36
.05	.83	.12	-.08	.29
.04	.83	.14	-.11	.21
.03	.82	.17	-.15	.21
.02	.81	.20	-.18	.19
.01	.79	.24	-.22	.19
.00	.78	.27	-.25	.17

* The average of the estimates of theta in the basic analysis is approximately 0.07.

This estimation process produced negative values of β for four of the classes, and of γ for one class. Such values are inconsistent with the definition of these parameters, and also can make it impossible to solve for θ in equation 9. Hence in these cases, the value 0 was substituted for the negative estimate, and the three equations solved approximately for the other two parameters.

The results are given in Table 12, which also contains the averages of the estimates of θ for the exercises taken by each class.

(2) Each subsequent week in which a class was tested, the learning parameter θ was estimated in the same way as in the basic analysis. These values are reported in Table 13 for two classes, the ones with the best, and worst overall predictions.

(3) Beginning with the 13th week, predictions were made each week from equation (3) for each class tested on that week, using the cumulative average of θ and the initial parameters for that class.

Table 14 contains a summary of the predictions for each class. The average estimates of θ range from -0.08 to 0.15, revealing a great difference among classes. Also, the predictions for each class are less accurate than those obtained for grand means. However, these results have been affected by the initial estimates of the skill parameters. Higher scores on comparable exercises were attained in weeks 11 and 12 than in later weeks, forcing early estimates of θ

Table 12

Estimates of Parameters for Individual Classes

Class	First Week	Beta	Mu	Gamma	Average Weekly Estimate of Theta
1	12	.00	.03	.22	-.03
2	12	.09	.00	.52	.03
3	11	.00	.08	.36	-.01
4	12	.02	.00	.27	.08
5	11	.04	.06	.13	-.05
6	11	.00	.07	.38	.03
7	12	.00	.13	.42	.11
8	11	.00	.03	.05	.15
9	11	.09	.00	.34	-.02
10	12	.07	.00	.68	.02
11	12	.08	.13	.00	.00
12	11	.00	.12	.10	.11
Average		.03	.05	.29	.02
Estimate Using Grand Means	11	.00	.09	.19	---
	12	.02	.07	.38	---

Table 13

Itemized Predictions for Two Classes

Class 9

Week	Exercise number	B	M	C	Theta Estimated	Proportion Correct Predicted	Observed
13	1	2	1	0	.00	.83	.76
13	2	1	1	0	.00	.91	.69
17	1	2	1	1	-.26	-.15	.87
17	2	2	2	1	-.26	-.15	.81
18	1	1	0	0	-.17	.73	.37
18	2	2	1	1	-.17	-.00	.56
18	3	4	1	1	-.17	-.17	.92
22	1	2	1	0	-.10	.54	.76
22	2	4	1	1	-.10	.00	.92
23	1	3	2	2	-.07	.03	.99
23	2	2	2	1	-.07	.15	.92
24	1	4	1	1	-.05	.16	.76
29	1	2	2	0	-.04	.68	.71
29	2	2	2	1	-.04	.24	.85

Note. The Average Absolute Value of Observed - Predicted = .52.

Class 6

Week	Exercise number	B	M	C	Theta Estimated	Predicted	Observed
13	1	2	1	0	.00	.93	.92
13	2	1	1	0	.00	.93	.92
18	1	1	0	0	-.04	1.00	.99
18	2	2	1	1	-.04	.47	.69
18	3	4	1	1	-.04	.47	.61
19	1	1	1	0	-.02	.92	.99
19	2	2	1	1	-.02	.52	.56
19	3	2	1	0	-.02	.92	.99
19	4	3	2	2	-.02	.27	.56
20	1	4	1	1	.02	.64	.78
20	2	3	2	0	.02	.89	.99
24	1	4	1	1	.05	.77	.50
25	1	2	1	1	.04	.75	.71
25	1	2	1	1	.04	.75	.71
25	2	2	2	1	.04	.71	.57
25	3	8	3	3	.04	.41	.57
26	1	1	0	0	.04	1.00	.88
26	2	4	1	1	.04	.75	.77

Note. The Average Absolute Value of Observed - Predicted = .11.

Table 14

Summary of Predictions for Each Class

Class	Average Estimated Theta	Average Absolute Difference	Average Difference	Proportion of Observed > Predicted
1	-.03	.39	-.33	.21
2	.03	.29	-.23	.28
3	-.01	.38	-.33	.19
4	-.08	.36	-.31	.25
5	-.05	.16	-.11	.33
6	.03	.11	-.04	.41
7	.11	.10	-.04	.40
8	.15	.14	-.01	.57
9	-.02	.52	-.47	.14
10	.02	.36	-.32	.21
11	.00	.19	-.12	.39
12	.11	.10	.05	.63
Average	.02	.26		

to reach high negative values. (Table 13 shows this is true even for the class with the most accurate predictions.)

The final evidence that weeks 11 and 12 provide poor initial estimates is that if the basic model, ignoring class membership, is used with initial parameters estimated from those weeks, then the average predictions are quite poor, the average absolute deviation being 0.52 (see Table 15). Thus the predictions by class, which at first seem so poor, may be better than those for the basic model, given the same set of exercises.

It may be recalled that week 11 produced the worst single prediction of the basic analysis (see Figure 5). However, the classes whose initial parameters were estimated on that week are not worse on the average than those estimated on week 12, which Figure 5 shows to be the most accurately predicted week of the year. Individual classes show too much variation in scores to produce accurate predictions from the model.

Three modifications of this analysis by class were carried out: using a constant learning parameter, using constant initial parameters across classes, and a combination of the two. The results are given in Table 15. Two values of the constant θ were used: 0.02, the average of the estimates for each class, and 0.07, the average obtained from the original analysis which ignored class. The table shows that using a constant learning parameter improves the estimation substantially and that it makes little difference whether the

Table 15

Summary of Results of Modified Class Analyses
(Average Absolute Deviations)

	Modification (see below)							
CLASS	A	B	B1	B2	C	D	D1	E
1	.39	.12	.17	.13	.18	.11	.11	---
2	.29	.21	.12	.26	.35	.13	.12	---
3	.38	.16	.14	.20	.37	.14	.14	---
4	.36	.08	.12	.09	.13	.08	.09	---
5	.16	.13	.20	.11	.27	.09	.16	---
6	.11	.08	.09	.11	.10	.08	.10	---
7	.10	.09	.11	.12	.18	.11	.16	---
8	.14	.11	.13	.11	.12	.17	.12	---
9	.52	.15	.11	.19	.76	.12	.12	---
10	.36	.19	.11	.25	.28	.14	.19	---
11	.19	.12	.16	.12	.39	.13	.15	---
12	.10	.08	.11	.10	.11	.12	.09	---
Average	.26	.13	.13	.15	.27	.12	.13	.52*

Key:

A: Basic analysis by class

B: Basic analysis with constant θ = 0.02 (the average of of estimates in (A)

B1: Same as (B) but with θ = 0.07 (the average of the basic model, ignoring class membership)

B2: Same as (B) but with θ = 0.00 (no learning)

C: Basic analysis with constant initial parameters - β = .03, μ = .05, γ = .29 (the averages of the estimates in (A)

D: Constant θ = 0.02 and initial parameters of type C

D1: Same as (D) but with θ = 0.07

E: Basic analysis ignoring class membership; initial parameters: β = .01, μ = .08, γ = .29 (the average of the estimates for the entire group from weeks 11 and 12

Note. All predictions begin with week 13.

* the original model gives a very high average error, but for other values of the initial parameters, this value is below 0.10.

same initial parameters are used for all the classes.

The result given in the table for the basic analysis (E) ignoring class, and beginning on week 13 is the most striking in Table 15. The value of the average absolute deviation (0.52), however, may be reduced to below the others in the table by merely adjusting the values of the initial parameters. That is, contrary to what we have seen for the basic analysis reported in Section IV-5, the estimates of the initial parameters made from weeks 11 and 12 are not nearly the best choices. For example, if $\beta = 0.1$, $\mu = 0.1$, and $\gamma = 0.3$, then the average absolute deviation becomes 0.09 for the basic model.

In summary, with this data, analysis by individual class raises more questions than it illuminates. The accuracy of the predictions for individual classes ranges very widely, the average absolute deviation having values between 0.10 and 0.52.

The erratic behavior of the predictions can be traced to apparently anomalous results on weeks 11 and 12, those chosen for the estimation of the initial skill parameters. The observed proportion correct on these weeks is substantially higher than on comparable exercises given in later weeks. This caused negative estimates of θ and produced very low (even negative!) predictions of proportion correct.

Any future investigation in this area should ensure a careful estimation of the initial parameters. On the other hand, the up and

down fluctuations of the observed proportion correct which is revealed by the analysis by individual class seems to be an undeniable attribute of this data. In this case, we must use caution in applying this learning theory model to smaller units of aggregation.

IV-7. Comparison with Regression Predictions

As discussed earlier, one alternative method for predicting proportion correct in the context of an elementary arithmetic curriculum is a regression of proportion correct on structural features of the exercises (see section III-6).

For purposes of comparison, a regression may be carried out using the values of b , m , and c from Table 4 as the independent variables and the observed proportion correct (also in Table 4) as the dependent variable. (Recall that b is the number of additions of two integers required, m is the number of moves to a new column, and c is the number of carries.)

As discussed in the earlier section, it is usual in regressions involving probabilities to transform the probabilities onto the set of real numbers. The transformation used here is $\log((1-p)/p)$. The procedure is as follows: Transform the dependent variable, perform the regression, and apply the inverse transformation to the regression estimates.

Table 16 shows the results of the regression on the transformed proportion correct. Note that only the coefficient of c , the number of carries, is statistically significant. The first two columns of Table 17 compare the predictions of the regression with those of the the basic model (Section IV-5). The results are extremely close, with the outstanding difference being that the correlation of observed and predicted is 0.80 with the model and 0.72 with regression.

Table 16

Regression of Observed Proportion Correct (P(TR))
on Structural Features of Exercises (B,M,C)

Sample Size: 42
Dependent Variable: P(TR)
Independent Variables: B M C

Coefficient of Determination: 0.522
Multiple Correlation Coefficient: 0.723

Estimated Constant Term: -1.604
Standard Error of Estimate: 0.570

Analysis of Variance for the Regression:

Source of Variation	DF	S. SQ.	M.S.	F	Prob.
Regression	3	13.478	4.493	13.84	0.000
Residuals	38	12.334	.325		
TOTAL	41	25.812			

Var.	Regression Coefficient	S. E. of Reg. Coef.	F-Value DF (1, 38)	Prob.	Corr. Coef. with P(TR)
B	-0.118	0.094	1.571	0.218	0.376
M	0.069	0.162	0.180	0.674	0.432
C	0.825	0.169	23.95	0.000	0.708

Table 17

Comparison of Predictions by Regression and Predictions by Model

	Model	Regression	'Predictive' Regression (see text)
Average Absolute Deviation	0.09	0.08	0.12
Correlation of Observed with Predicted	0.80	0.72	0.46
Average Deviation	-0.02	0.01	0.01
Proportion of Predictions > Observed	0.48	0.52	0.52

It must also be pointed out again that the model predictions are obtained without using the respective observed values, while in this case the regression analysis uses all the data to obtain a prediction.

Figures 7 and 8 are graphs of predicted vs. observed and residuals vs. week, which are to be compared with Figures 4 and 5 of Section IV-5. Again, there is no great difference among the figures.

A final comparison may be made between the regression estimates of proportion correct and those of the model. To test the predictive power of the regression a subset of the exercises may be omitted from the regression, and then the regression equation used to predict the scores on these items. The subsets to be omitted may be chosen in various ways; the approach used here was to omit the items in each exercise class, one set at a time. That is, predictions were made for each exercise from the regression equation obtained by omitting all exercises of the same type.

The results of these predictions are given in the third column of the table. When used in a truly predictive manner, the regression approach provides predictions whose deviations are slightly higher than those of the model, and with a correlation that is substantially lower.

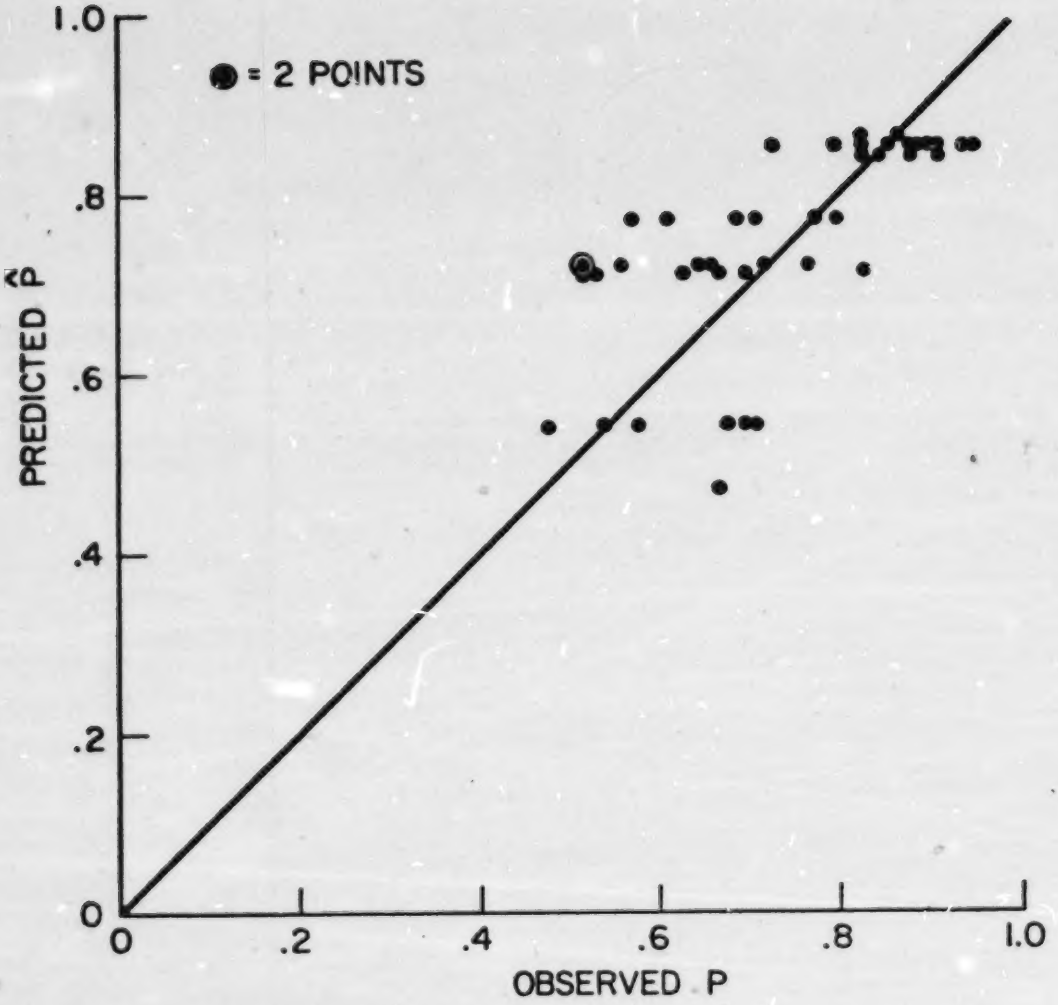


FIGURE 7: REGRESSION PREDICTIONS VS. OBSERVED PROPORTION CORRECT

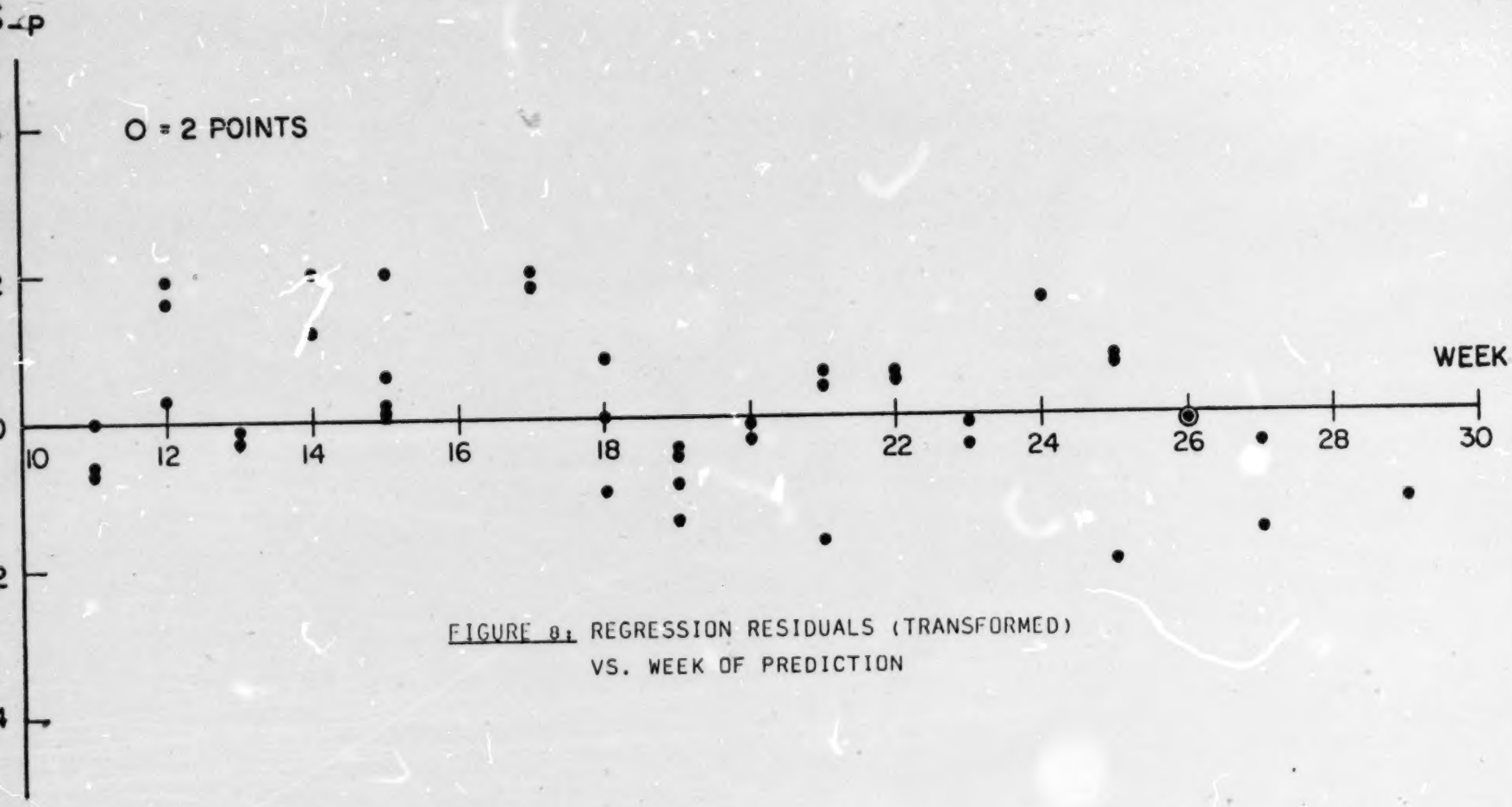


FIGURE 8: REGRESSION RESIDUALS (TRANSFORMED) VS. WEEK OF PREDICTION

V. SUMMARY AND DISCUSSION

Table 18 gives a summary of the accuracy of the various predictions described in the previous sections. The salient features are as follows:

(1) The basic model predicts proportion correct on exercises with an overall average absolute deviation of about 0.09 with a correlation between observed and predicted scores of 0.80. This is remarkably accurate when viewed in the context of the results of Suppes and Morningstar (1972). In that study one set of 20 vertical addition items is predicted with an average absolute deviation of 0.05, but the prediction process incorporated the values to be predicted, which was not the case in the present study.

The predictions may also be compared with those obtained from regression on structural features by Suppes, Jerman, and Hyman (1967). It may be calculated from the representative subset of addition exercises which are reported in detail (p.182) that the average absolute deviation between predicted and observed proportion correct is 0.12, and the correlation between observed and predicted is 0.82.

(2) The initial skill levels estimated on week 9 are in fact the ones which produce the best predictions on the average.

(3) The average value of the learning parameter θ is 0.07, which is comparable to the values found in the literature for simple learning (Coombs, Dawes, and Tversky, 1970, p. 264)

Table 18

Summary of Results

DESCRIPTION	SECTION IN TEXT	AVERAGE ABSOLUTE DEVIATION (Correlation in parentheses) MODEL*	LOWEST**
BASIC MODEL - IGNORING CLASS MEMBERSHIP (First Week Predicted: 10)			
Basic Analysis	IV-5:A	.09 (.80)	.09 (.80)
Reduced Models	IV-6:A		
2 parameter: beta,mu		.23 (.01)	.11 (.43)
beta,gamma		.09 (.80)	.09 (.80)
mu,gamma		.13 (.85)	.10 (.85)
1 parameter: beta			.14 (.28)
mu			.11 (.25)
gamma			.12 (.81)
Constant θ	IV-6:B		
θ = .00 (no learning)		.27 (.78)	---
θ = .07 (average value)		.09 (.83)	---
Regression on structural features	IV-7		
Standard		.08 (.72)	---
"Predictive"		.12 (.46)	---
ANALYSIS BY CLASS (First Week Predicted: 13)			
Basic Analysis	IV-6:C	.26	---
Constant learning parameter			
θ = .02 (average)		.13	---
θ = .00 (no learning)		.15	---
θ = .07 (average of original)		.13	---
Constant initial parameters		.27	.14
Ignoring Class (for comparison)	IV-6:C		
initial parameters estimated on weeks 11 and 12		.52	.09
initial parameters estimated on week 9		.17	---

* Initial skill parameters estimated as prescribed by the model

** Initial skill parameters varied to find lowest value of average absolute deviation

(4) Reduced models have varying predictive success. If the parameter δ (probability of correctly carrying to the next column) is included, then two parameter and one parameter models have overall accuracy quite close to the full model.

(5) When the model is used to predict proportion correct for an exercise to be tested some weeks hence, the predictions are stable for about ten weeks prior to the week of testing. Thus the predictions of the score on an exercise taken on week n are little different if made on week $n - 10$ or on week $n-1$.

(6) The linear learning relationship which has been used assumes a constant value of the learning parameter θ, and when this is used throughout, the predictions have an overall accuracy comparable to that of the basic analysis, in which θ is estimated each week. If the value 0.00 is used for the constant learning parameter, this is analogous to assuming no learning, and the results show that the model then becomes much less accurate in predicting future scores.

(7) Regression on structural features (and ignoring any learning) produces results which are not much different from the basic model, although the correlation is somewhat lower. When all the exercises of the same type are omitted in turn from the regression, and then the scores on these exercises predicted, the regression deviations rise to 0.12 and the correlation falls to 0.46.

(8) Comparing the analysis by individual class with the basic analysis, which ignores class membership, is a complex matter. For one reason, the initial skill parameters estimated on weeks 11 and 12 produce poorer predictions than when other values are used. Also, there is a great deal of variation of individual class scores on comparable exercises over time, and although some of the predictions made for individual classes are as accurate as those for the entire group (see Table 15), for several of them the predictions of the model are consistently poor.

In summary, this model incorporating learning with structural features of vertical addition exercises predicts proportion correct with an accuracy comparable to other methods which have been employed in the literature. This study indicates that the most important skill to consider in making these predictions is carrying. For these second-grade students, the skill of adding two digits makes a much smaller contribution. Finally, using classes as the unit of analysis does not appear to increase the overall accuracy of the predictions.

VI. Suggestions for Future Research

This study was carried out using tests which were designed for use in formative evaluation of the Radio Mathematics curriculum in Nicaragua. It is obvious that similar research with data from other countries and other arithmetic curricula is necessary. This section will emphasize instead emphasize the differences in design which should be considered.

(1) More exercises per week. This would not only provide more accurate estimates of the parameters of the model, but it would allow the introduction of other parameters. An important addition would be to introduce a learning parameter for each of the three sub-operations so that it would not be necessary to assume that skill levels in all of these areas increase at the same rate.

The learning parameter θ is estimated once for each exercise on a given week, and then averaged across exercises. To estimate three learning parameters would require a minimum of three exercises, and to sharpen the estimate by using an average across exercises would require some multiple of three exercises.

More exercises per week would also allow the repetition of the same exercise type each week, and then it would be possible to begin to study the learning by individuals. A study could be planned in which each student in a subset of the sample would receive say ten

exercises of each of four types of exercises. We would estimate the individual's probability of a correct response using the observed proportion correct on the problems of each type.

(2) Greater variety of exercises. The list of exercises used in this study (Tables 2 and 4) reveal that the variety of exercises is limited. There are few problems involving three digit numbers and few with more than two addends. Thus the relationship $b = m + 1$ is commonly encountered, and it is difficult to distinguish between the contribution of these sub-operations.

With more different kinds of exercises it would also be possible to study in greater detail the influence of exercise type on accuracy of predictions.

(3) Initial Parameters. In this study it was shown that when the class membership is ignored, the values of the initial parameters as estimated from an early test week proved to produce more accurate predictions than any other set of values of these parameters. This was a very important result and should be re-verified.

However the picture was much less clear when class membership was taken into account in the analysis. Any future study will have to attend carefully to the estimation of these parameters. this would be aided by the inclusion of at least six exercises of varied types on each of three early weeks.

(4) The class analysis. If individual differences are to be

studied, then classes must be the first step. The class analysis in this study is perhaps its most unsatisfactory aspect. The confusion in results may be intrinsic, or it may be clarified by a better design. The design could be improved by: testing all classes each week; by taking some preliminary measures in each class to determine what social, environmental, or achievement differences exist among them; or by controlling the manner of presentation of the tests to minimize these differences across classes.

Furthermore, classes which have missed some teaching on a given week, or whose day or week is somehow disrupted should not be included on that week, or should be eliminated altogether. Unexpected fluctuations, as were observed in the present study, should be discussed with the teacher and/or observers, and an explanation sought, for scores which rise and fall erratically are direct evidence against some of the fundamental assumptions of the learning model.

(5) Variations in the model. The above suggestions might lead to a more reliable, valid, and defensible investigation of the fit of this learning theory/structural analysis model to the results of classroom tests involving vertical addition exercises. But the model could be modified in other basic ways. For example, the assumption of independence could be tested by assuming some dependence among the three sub-operations or among sequential applications of the same sub-operation. This would require the estimation of some conditional probabilities, and would necessitate the inclusion of more exercises per week.

The study could include more or different sub-operations. Suppes and Morningstar (1972) and others who have analyzed errors in this type of exercise include many suggestions for sub-operations (sources of error): entering the number correctly given that the addition was done correctly; entering the column sum in the correct position; or mentally storing the correct cumulative column sum in a problem with more than two addends.

The distinction could also be made between adding two digits and adding a digit to a number greater than 10, when necessary within a column in an exercise with more than two addends. Or the sub-operation of carrying could be divided into two parts: writing down the correct digit and adding the correct digit to the next column.

Each of these suggested additions would enrich the model and require additional exercises for the estimation of the additional parameter. However, the study could perhaps then point to a subset sub-operations which economically predict proportion correct.

(6) Other kinds of exercises. Finally, the study could be extended to other kinds of addition exercises (presented orally, in a horizontal format, or presented in text). And as Suppes and Morningstar have done, a study of other operations, in particular subtraction and multiplication could be carried out.

VII. REFERENCES

Bruner, J. S., "Some theorems on instruction illustrated with reference to mathematics", in N.S.S.E. 63rd Yearbook, Theories of Learning and Instruction. Chicago: The University of Chicago Press, 1964

Bush, R. R., "A survey of mathematical learning theory", in Luce, D. R. (ed.), Developments in Mathematical Psychology. The Free Press of Glencoe, Illinois, 1960

Bush, R. R. and Mosteller, F., Stochastic Models of Learning. New York: John Wiley and Sons, 1955

Coombs, C. H., Dawes, R. M., Tversky, A., "Mathematical learning theory", Mathematical Psychology (Chapter 9). Prentice Hall, Englewood Cliffs, 1970

Cotton, J. W., "Models of learning", Annual Review of Psychology, Volume 27, p. 155-185, 1976

Cromer, F.E., "Structural models for predicting the difficulty of multiplication problems", Journal for Research in Mathematics Education, Vol. 5, p. 155-166, 1974

Cronbach, L.J., "Research on classrooms and schools: Formulation of questions, design, and analysis", Occasional paper of the Stanford Evaluation Consortium, Stanford University, 1976

Cronbach, L. J. and Suppes, P. C., "Research For Tomorrow's Schools". New York: MacMillan, 1969

Estes, W. K., "Toward a statistical theory of learning" Psych. Review 1950, 57, p. 94-107

Estes, W. K., "Learning Theory". Annual Review of Psychology, 1962, 13, p. 107-144.

Estes, W. K. and Suppes, P., "Foundations of linear models" in Bush, R. R. and Estes, W. K., (Eds.), Studies in Mathematical Learning Theory. Stanford University Press, 1959

Estes, W. K. and Suppes, P., "Foundations of Stimulus Sampling Theory". In Krantz, D. H., Atkinson, R. C., Luce, R. D., and Suppes, P. (Eds.), Contemporary Developments in Mathematical Psychology, Volume 1: "Learning, Memory, and Thinking". San Francisco: W. H. Freeman, 1974.

Greeno, J. G., and Bjork, R. A., "Mathematical learning theory and the new 'mental forestry' ". Annual Review of Psychology, Vol. 24, p. 81-116, 1973

Joncich, G., "The Sane Positivist: A Biography of E. L. Thorndike". Middletown, Connecticut: Wesleyan University Press, 1968.

Luce, R. D., Individual Choice Behavior. New York: Wiley, 1959

Searle,B., Friend, J., and Suppes, P. The Radio Mathematics Project: Nicaragua, 1974-75. Stanford, Calif.: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1976

Searle, B., Matthews, P., Suppes, P., and Friend, J., "Formal Evaluation of the Radio Mathematics Instructional Program, Nicaragua - Grade 1, 1976." In Cook, T., Editor, Evaluation Studies Review Annual, Volume 3. Sage Publications, 1978.

Suppes, P., and Atkinson, R. Markov learning models for multiperson interactions. Stanford, California: Stanford Univ. Press, 1960

Suppes, P., "Facts and fantasies in education," Phi Delta Kappa monograph series. Bloomington, Indiana: 1973

Suppes, P., Hyman, L., and Jerman, M., "Linear structural models for response and latency performance in arithmetic on computer controlled terminals." In J.P. Hill (Ed.), Minnesota symposia on child psychology. Minneapolis: University of Minnesota Press, 1967.

Suppes, P., and Morningstar, M., Computer-assisted instruction at Stanford, 1966-68: Data, models, and evaluation of the arithmetic programs. New York: Academic Press, 1972

Thorndike, E. L., "The Psychology of Arithmetic". New York: Mac Millan, 1922.

Wagner, W. J., "Thorndike and Arithmetic: A study of the influence of research on teaching." Unpublished manuscript; Stanford University, 1975.

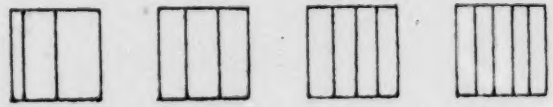
APPENDIX A - Sample Student Answer Sheets

Three forms of the test given on week 15 are included here. The vertical addition items are C (AV19), E (AV01), H (AV12), and K (AV27). The different tests sheets were distributed at random within each class.

It may be seen from these sample tests that there are two different modes of presentation - oral and written. Exercises A, G, and L on the sheets are presented by the teacher orally and have only a line for the student to write a response. In these cases, of course, all children in each classroom are given the same form of the exercise. Three different forms of these exercises do exist, however, and are assigned to classes rather than to children within classes.

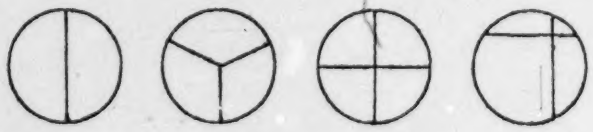
All the vertical addition exercises studied in this dissertation were presented in the manner depicted on these sample answer sheets.

Nombre

A _____	**G** _____
B 33 23 13	**H** 246 + 696
C 5 + 8	**I** 29 − 23
D 154	**J** 
E 0 + 8	**K** 28 18 + 23
F 2 x 3 = _____	**L** _____

2-15

Nombre

A _____	**G** _____
B 35 25 52	**H** 156 +177
C 7 +4	**I** 75 -72
D 243	**J** 
E 5 +3	**K** 15 60 +17
F 5 x 0 = _____	**L** _____

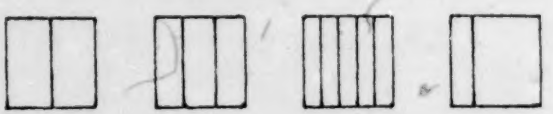
2-15

126

b.

127ii

Nombre

A _____	**G** _____
B 74 64 47	**H** 161 + 679
C 8 + 9	**I** 56 − 50
D 512	**J** 
E 2 + 7	**K** 35 13 + 4
F 1 x 1 = _____	**L** _____

2-15

APPENDIX B - The Likelihood Estimate for a Week

For w exercises on week $j+1$, we have the following expression for the likelihood L (the probability of obtaining the particular observations encountered):

$$L = \prod_{i=1}^{w} \left(B^{b_i} M^{m_i} C^{c_i}\right)^{n_i} \left(1 - B^{b_i} M^{m_i} C^{c_i}\right)^{N-n_i}$$

where $B = 1-(1-\theta)^j \beta$

$M = 1-(1-\theta)^j \mu$

$C = 1-(1-\theta)^j \gamma$

b_i, m_i, c_i = the number of the respective sub-operations required by exercise i.

N = the number of students responding (constant each week)

n_i = the number of students correctly answering exercise i.

The likelihood may also be written:

$$L = B^{\sum b_i n_i} M^{\sum m_i n_i} C^{\sum c_i n_i} \prod_{i=1}^{w} \left(1 - B^{b_i} M^{m_i} C^{c_i}\right)^{N-n_i}$$

Letting $b = \sum b_i n_i$, $c = \sum c_i n_i$, and

$m = \sum m_i n_i$

and taking the logarithm, we have:

$$\ell = \log L = b \log B + m \log M + c \log C + \sum [(N - n_i) \log (1 - B^{b_i} M^{m_i} C^{c_i})]$$

Taking the derivative with respect to θ,

$$\frac{\partial \ell}{\partial \theta} = j(1-\theta)^{j-1} \left[\frac{b\beta}{B} + \frac{m\mu}{M} + \frac{c\gamma}{C} \right] - j(1-\theta)^{j-1} \sum \left(\frac{N - n_i}{1 - p_i} p_i K_i \right)$$

where $p_i = 1 - B^{b_i} M^{m_i} C^{c_i}$ = predicted proportion correct for exercise i

and $K_i = \frac{b_i \beta}{B} + \frac{m_i \mu}{M} + \frac{c_i \gamma}{C}$

Since $\sum n_i K_i = \frac{b\beta}{B} + \frac{c\gamma}{C} + \frac{m\mu}{M}$

then $\frac{\partial \ell}{\partial \theta} = j(1-\theta)^{j-1} \left[\sum m_i K_i - (N - m_i) \frac{p_i}{1 - p_i} K_i \right]$

$$= j(1-\theta)^{j-1} \left[\sum \left(\frac{K_i}{1 - p_i} (m_i - N p_i) \right) \right]$$

An obvious solution of $\frac{\partial \ell}{\partial \theta} = 0$ is $\frac{m_i}{N} = p_i$ for each i, but this gives a different value of θ for each exercise. The unique solution for the value of θ for which $\frac{\partial \ell}{\partial \theta} = 0$, however, is much more difficult to obtain.